

---

# Masterarbeit

---

Herr  
**Eric Frenzel**

**Erneuerbare Energie - Darstellung  
des regulativen Netzwerks der bio-  
logischen Wasserstoffproduktion  
mittels bioinformatischer Visuali-  
sierungsmethoden**

**2013**

# **Masterarbeit**

---

## **Erneuerbare Energie - Darstellung des regulativen Netzwerks der biologischen Wasserstoffproduktion mittels bioinfor- matischer Visualisierungsmethoden**

Autor:

**Herr**

**Eric Frenzel**

Studiengang:

**Molekularbiologie/Bioinformatik**

Seminargruppe:

**MO10w1-M**

Erstbetreuer:

**Herr**

**Prof. Dr. rer. nat. habil. Röbbbe Wünschiers**

Zweitbetreuer:

**Herr**

**Prof. Dr. rer. nat. Dirk Labudde**

---

## Danksagung

Mein Dank gebührt an erster Stelle meiner Familie, die mich in Zeiten von Bergen und Tälern mit Erfolg und Rückschlägen über meine gesamte Studienzeit stets ermutigt, unterstützt und gefördert hat.

Großen Dank verpflichtet bin ich meinem Betreuer *Prof. Dr. rer. nat. habil.* Röb-  
be Wünschiers für die Ermöglichung dieser Arbeit unter seiner Leitung. In die-  
ser Zeit bin ich nicht nur als Wissenschaftler, sondern auch als Mensch gereift.

Weiterhin danke ich Jonathan Ross Karr, Doktorand der Biophysik, von der  
Stanford University für seine Unterstützung beziehungsweise der weiterführenden  
Etablierung einer *Knowledgebase* für Cyanobakterien.

Mein abschließender Dank gilt unterstützenden Freunden/-innen sowie der  
Fachgruppe Biotechnologie der Hochschule Mittweida für ihre Tipps und Rat-  
schläge in schwierigen Situationen.

---

## **Bibliografische Angaben**

Frenzel, Eric

Erneuerbare Energie - Darstellung des regulativen Netzwerks der biologischen Wasserstoffproduktion mittels bioinformatischer Visualisierungsmethoden

68 Seiten, Hochschule Mittweida - University of Applied Sciences, Fakultät Mathematik/Naturwissenschaften/Informatik, Masterarbeit, 2013

---

## Referat

Cyanobakterien gehören zu den evolutionär ältesten Bewohnern unserer Erde und haben sich während Jahrmilliarden an stark differierende Umweltbedingungen adaptiert. Sie besiedeln teils Habitate mit widrigsten Konditionen, seien es bspw. im höchsten Maß nährstoffarme Gebiete oder jene mit extremer Hitze [Fuchs & Schlegel, 2007]. Durch die genomischen Beschaffenheiten sind Cyanobakterien in der Lage, Wasserstoff zu produzieren. Durch bioverfahrenstechnische Fortschritte sowie den biologisch-fachlichen Fortschritt der Kenntnisse der Stoffwechselvorgänge kann es in einigen Jahren vermutlich möglich werden, dass auf effektive Weise dieser Biowasserstoff gewonnen und als regenerativer Kraftstoff eingesetzt werden kann [URL-4; URL-5; URL-6; Bandyopadhyay *et al.*, 2010]. Dadurch könnte in einem hohen Maß der Einsatz an fossilen Energieträgern reduziert werden. Im Umfang der vorliegenden Masterarbeit werden theoretische sowie experimentelle metabolische Daten des aktuellen wissenschaftlichen Standes der Cyanobakterien *Synechocystis* sp. PCC 6803 und *Nostoc* sp. PCC 7120 mittels einer umfangreichen Literaturrecherche aufgearbeitet, eine Analyse von Visualisierungssoftware durchgeführt und metabolische Modelle des Wasserstoffmetabolismus erstellt. Weiterhin wird die Thematik der Etablierung einer *Knowledgebase* für Cyanobakterien diskutiert [Karr *et al.*, 2012].

Diese Arbeit ist Teil des EU-Forschungsprojekts CyanoFactory (FP 7, Fördernummer 308518).

---

## Abstract

Cyanobacteria are among the evolutionary oldest organisms on earth. They have been evolving since billions of years and are adapted to different environmental conditions [Fuchs & Schlegel, 2007]. Because of their genomic texture, cyanobacteria have the possibility to produce hydrogen gas. As a result of progressive development in biotechnological and procedural knowledge of their metabolic pathways, it is possible that in some years biohydrogen will be in the lead as a regenerative fuel [URL-4; URL-5; URL-6; Bandyopadhyay *et al.*, 2010]. Thus, the usage of fossil fuel can be strongly reduced. For this master thesis, theoretical and experimental data of the cyanobacteria *Synechocystis* sp. PCC 6803 and *Nostoc* sp. PCC 7120 have been collected and used to establish a metabolic and regulatory network of hydrogen metabolism. To achieve that, different visualization software for biological processes and metabolic pathway drawing have been tested. Furthermore, the establishment of a *Knowledgebase* for cyanobacteria, similar to that of *Mycoplasma genitalium*, is discussed [Karr *et al.*, 2012].

This thesis is part of the EU-funded research project CyanoFactory (FP 7, Energy Contract 308518).

# INHALTSVERZEICHNIS

INHALTSVERZEICHNIS.....	I
ABKÜRZUNGSVERZEICHNIS .....	III
ABBILDUNGSVERZEICHNIS.....	IV
TABELLENVERZEICHNIS.....	VI
ZIELSTELLUNG.....	VII
<b>1 BIOLOGISCHE GRUNDLAGEN .....</b>	<b>1</b>
1.1 CYANOBAKTERIEN .....	1
1.2 WASSERSTOFFPRODUKTION IN CYANOBAKTERIEN .....	2
1.2.1 [NiFe]-HYDROGENASEN .....	3
1.2.2 NITROGENASE .....	8
1.3 UNTERSUCHUNGSORGANISMEN .....	10
1.3.1 <i>SYNECHOCYSTIS</i> SP. PCC 6803.....	11
1.3.2 <i>NOSTOC</i> SP. PCC 7120 .....	17
<b>2 MATERIAL UND METHODEN .....</b>	<b>26</b>
2.1 SOFTWARE.....	26
2.1.1 BIOINFORMATISCHE VISUALISIERUNGSSOFTWARE .....	26
2.1.2 VANTED .....	28
2.2 KONZEPTIONELLE VISUALISIERUNG .....	30
<i>SYSTEMS BIOLOGY GRAPHICAL NOTATION</i> .....	30
2.3 DATENBANKEN .....	34
2.3.1 PROMBASE .....	34
2.3.2 KEGG .....	35
2.3.3 BiOCYC .....	36
2.3.4 SYNECHONET.....	36
2.3.5 STRING .....	37
2.3.6 ODB.....	38
2.3.7 OPERONDB.....	38
2.3.8 GENOMICS10 .....	39
2.4 <i>KNOWLEDGEBASE</i> .....	39
<i>KNOWLEDGEBASE VON MYCOPLASMA GENITALIUM ALS VORLAGE FÜR</i> CYANOBAKTERIEN .....	39
2.5 STUDIE.....	41
VISUALISIERUNGSSTANDARDS IN DER BIOLOGIE.....	41

---

<b>3</b>	<b>ERGEBNISSE UND DISKUSSION .....</b>	<b>43</b>
3.1	VISUALISIERUNG METABOLISCHER ABLÄUFE MIT VANTED SOWIE DER <i>SYSTEMS BIOLOGY GRAPHICAL NOTATION</i> .....	43
3.1.1	METABOLISCHE KARTE VON <i>SYNECHOCYSTIS</i> SP. PCC 6803 .....	43
3.1.2	METABOLISCHE KARTE VON <i>NOSTOC</i> SP. PCC 7120 .....	47
3.2	DETEKTION VON OPERONS .....	48
3.3	PROTEIN-PROTEIN-INTERAKTIONEN .....	52
3.4	<i>KNOWLEDGE</i> BASE FÜR CYANOBAKTERIEN .....	55
3.5	VERGLEICH VON VISUALISIERUNGS SOFTWARE FÜR DIE DARSTELLUNG BIOLOGISCHER UND METABOLISCHER ABLÄUFE .....	58
3.6	AUSWERTUNG DER STUDIE ÜBER DIE AKZEPTANZ EINES VISUALISIERUNGSSTANDARDS IN DER BIOLOGIE .....	61
<b>4</b>	<b>ZUSAMMENFASSUNG .....</b>	<b>66</b>
<b>5</b>	<b>AUSBLICK .....</b>	<b>67</b>
	<b>LITERATURVERZEICHNIS.....</b>	<b>VII</b>
	<b>ANLAGEN .....</b>	<b>XIX</b>
	<b>EIGENSTÄNDIGKEITSERKLÄRUNG .....</b>	<b>XXV</b>



## ABKÜRZUNGSVERZEICHNIS

<b>API</b>	Programmierschnittstelle
<b>bp</b>	<b>B</b> asen <b>p</b> aar(e)
<b>CN</b>	Cyanid-Gruppe
<b>CO</b>	Kohlenstoffmonoxid
<b>CO<sub>2</sub></b>	Kohlenstoffdioxid
<b>CoA</b>	<b>C</b> oenzym <b>A</b>
<b><i>E. coli</i></b>	<b><i>Escherichia coli</i></b>
<b>EBI</b>	<b>E</b> uropean <b>B</b> ioinformatics <b>I</b> nstitute
<b>Fe</b>	Eisen
<b>hox</b>	hydrogen <b>o</b> xidation
<b>hup</b>	hydrogenase <b>u</b> ptake
<b>hyp</b>	hydrogenase <b>p</b> leiotropic
<b>kb</b>	<b>K</b> ilo- <b>B</b> asenpaare
<b>kDa</b>	<b>K</b> ilodalton
<b>KEGG</b>	<b>K</b> yoto <b>E</b> ncyclopedia of <b>G</b> enes and <b>G</b> enomes
<b>MFA</b>	<b>m</b> etabolic <b>f</b> lux <b>a</b> nalysis
<b>Mo</b>	Molybdän
<b>Ni</b>	Nickel
<b>nif</b>	<b>n</b> itrogen <b>f</b> ixation
<b>nt</b>	<b>N</b> ukleotide
<b>O<sub>2</sub></b>	Sauerstoff
<b>ODB</b>	<b>O</b> peron <b>D</b> ata <b>B</b> ase
<b>ORF</b>	<b>O</b> pen <b>R</b> eadin <b>G</b> <b>F</b> rame
<b>PCC</b>	<b>P</b> asteur <b>C</b> ulture <b>C</b> ollection of Cyanobacteria
<b>PPI</b>	<b>P</b> rotein- <b>P</b> rotein- <b>I</b> nteraktion
<b>SBGN</b>	<b>S</b> ystems <b>B</b> iology <b>G</b> raphical <b>N</b> otation
<b>SOM</b>	<b>S</b> elf <b>O</b> rganizing <b>M</b> ap
<b>STRING</b>	<b>S</b> earch <b>T</b> ool for the <b>R</b> etrieval of <b>I</b> nteracting <b>G</b> enes
<b>URI</b>	<b>U</b> niform <b>R</b> esource <b>I</b> dentifier
<b>v. l. n. r.</b>	von links nach rechts
<b>XML</b>	<b>E</b> xtensible <b>M</b> arkup <b>L</b> anguage

## ABBILDUNGSVERZEICHNIS

Abbildung 1: Schematische Darstellung des aktiven Zentrums einer [NiFe]-Hydrogenase (nach Hoffmann, 2008).....	4
Abbildung 2: Die bidirektionale Hydrogenase als schematische Darstellung (nach Tamagnini <i>et al.</i> , 2007) .....	5
Abbildung 3: Die uptake Hydrogenase als schematische Darstellung (nach Tamagnini <i>et al.</i> , 2007) .....	7
Abbildung 4: Die Nitrogenase als schematische Darstellung (nach Tamagnini <i>et al.</i> , 2007).....	10
Abbildung 5: Rasterelektronenmikroskopische Aufnahme von <i>Synechocystis</i> sp. PCC 6803 (nach van de Meene, 2005) .....	11
Abbildung 6: Metabolische Karte für <i>Synechocystis</i> sp. PCC 6803 nach Navarro <i>et al.</i> (2009) .....	14
Abbildung 7: <i>hox</i> -Operon von <i>Synechocystis</i> sp. PCC 6803 (nach Hoffmann, 2008).....	16
Abbildung 8: Mikroskopische Aufnahme von <i>Nostoc</i> sp. PCC 7120 Filamenten [Florres & Herrero, 2010] .....	19
Abbildung 9: Schematische Darstellung der chromosomalen Abfolge im Bereich der umstrukturierten DNA-Elemente <i>nifD</i> und <i>fdxN</i> [Carrasco <i>et al.</i> , 1994].	25
Abbildung 10: Schematische Darstellung der chromosomalen Abfolge im Bereich des umstrukturierten DNA-Elements <i>hupL</i> durch die Rekombinase <i>xisC</i> (nach Carrasco <i>et al.</i> , 1995).....	25
Abbildung 11: Flussdiagramm - Arbeitsweise mit der Software VANTED (nach Junker <i>et al.</i> , 2006) .....	29
Abbildung 12: Übersicht über die Entity Pool Nodes der Process description language .....	32
Abbildung 13: Übersicht über die Process Nodes der Process description language .....	32
Abbildung 14: Übersicht über die Container Nodes der Process description language .....	32
Abbildung 15: Übersicht über die Reference Nodes der Process description language .....	33
Abbildung 16: Übersicht über die Connecting Arcs der Process description language .....	33
Abbildung 17: Übersicht über die Logical Operators der Process description language .....	33
Abbildung 18: Übersicht über die Auxiliary Units der Process description language .....	34
Abbildung 19: Ausschnitt eines mit VANTED und des SBGN-Standards erstelltes metabolischen Netzwerk .....	46
Abbildung 20: Flussdiagramm - interaktive Verlinkung eines ORF im metabolischen Modell zu dem zugehörigen Protein-Protein-Interaktionsnetzwerk in der STRING-Datenbank .....	54
Abbildung 21: Zugriff auf hinterlegte Links in VANTED .....	55
Abbildung 22: Arbeitsoberfläche Knowledgebase .....	56

---

<b>Abbildung 23: Verteilung der Antworten auf die Frage „Finden Sie die Etablierung eines Visualisierungsstandards in der Biologie sinnvoll?“ .....</b>	<b>62</b>
<b>Abbildung 24: Verteilung der Antworten auf die Frage „Haben Sie bereits von dem Visualisierungsstandard SBGN gelesen oder sogar angewendet?“ ....</b>	<b>63</b>
<b>Abbildung 25: Verteilung der Antworten auf die Frage „Können Sie sich vorstellen, SBGN in Zukunft/weiterhin für Ihre Visualisierung anzuwenden?“ .....</b>	<b>64</b>

## TABELLENVERZEICHNIS

Tabelle 1: Auflistung der ORFs und zugehöriger Gen-Namen der bidirektionalen Hydrogenase-Untereinheiten sowie der akzessorischen Gene von <i>Synechocystis</i> sp. PCC 6803.....	16
Tabelle 2: Auflistung der ORFs und zugehöriger Gen-Namen der Nitrogenase von <i>Nostoc</i> sp. PCC 7120 .....	20
Tabelle 3: Auflistung der ORFs und zugehöriger Gen-Namen der akzessorischen Gene von <i>Nostoc</i> sp. PCC 7120, die für die Reifung der bidirektionalen Hydrogenase entscheidend sind .....	21
Tabelle 4: Auflistung der ORFs und zugehöriger Gen-Namen der bidirektionalen Hydrogenase-Untereinheiten von <i>Nostoc</i> sp. PCC 7120 .....	22
Tabelle 5: Auflistung der ORFs und zugehöriger Gen-Namen der uptake Hydrogenase-Untereinheiten von <i>Nostoc</i> sp. PCC 7120 .....	23
Tabelle 6: Auflistung der DNA-Elemente, die während der Heterocysten-Ausbildung verändert werden, die daran beteiligten Rekombinasen sowie die Beschreibung der genomischen Umordnung [Ramaswamy <i>et al.</i> , 1997; Carrasco <i>et al.</i> , 2005; Wei <i>et al.</i> , 1994] .....	24
Tabelle 7: Auflistung der getesteten Softwaretools für die Visualisierung biologischer Netzwerke mit Angabe der Versionsnummer sowie dem Betriebssystem, auf dem die Software installiert und getestet wurde .....	27
Tabelle 8: Darstellung der identifizierten Operons von <i>Synechocystis</i> sp. PCC 6803 im betrachteten genomischen Ausschnitt mittels der drei Datenbanken Genomics10, OperonDB und ODB .....	49
Tabelle 9: Darstellung der identifizierten Operons von <i>Nostoc</i> sp. PCC 7120 im betrachteten genomischen Ausschnitt mittels der drei Datenbanken Genomics10, OperonDB und ODB .....	51
Tabelle 10: Übersicht der Möglichkeiten getesteter Visualisierungsprogramme .....	60
Tabelle 11: Quantitative Auflistung des Teilnahmeergebnisses der Studie über die Evaluierung der Akzeptanz eines Visualisierungsstandards in der Biologie .....	62
Tabelle 12: Darstellung einer Auswahl von Antworten auf die Frage „Stellen Sie explizite Ansprüche an einen Visualisierungsstandard? Gern können Sie nachfolgend einige Ihrer Wünsche äußern.“ .....	65
Tabelle 13: Datensatz im metabolischen Modell involvierter Proteine und die zugehörigen ermittelten Promoter-Bindestellen und deren Reliabilität mittels PromBase von <i>Synechocystis</i> sp. PCC 6803.....	XIX
Tabelle 14: Datensatz im metabolischen Modell involvierter Proteine und die zugehörigen ermittelten Promoter-Bindestellen und deren Reliabilität mittels PromBase von <i>Nostoc</i> sp. PCC 7120 .....	XXI
Tabelle 15: Darstellung der weiteren Antworten bezüglich der Ansprüche an einen Visualisierungsstandard .....	XXIV

## ZIELSTELLUNG

Die vorliegende Arbeit ist im Rahmen des Masterstudiums Molekularbiologie/Bioinformatik an der HS Mittweida angefertigt worden und beschreibt das systematische Vorgehen einer Darstellung des regulativen Netzwerks der biologischen Wasserstoffproduktion mittels bioinformatischer Visualisierungsmethoden in Cyanobakterien. Als Vorhaben ist thematisiert worden, eine umfassende Literaturrecherche und die konsekutive kompakte Formulierung wissenschaftlicher Grundlagen und Erkenntnisse auf dem Gebiet der Cyanobakterien sowie des Wasserstoffmetabolismus dieser Prokaryoten durchzuführen. Außerdem sollen mit Hilfe geeigneter bioinformatischer Software die metabolischen Abläufe der biologischen Wasserstoffproduktion in den Cyanobakterien *Synechocystis* sp. PCC 6803 und *Nostoc* sp. PCC 7120 visualisiert werden.

# 1 BIOLOGISCHE GRUNDLAGEN

Die biologischen Grundlagen bilden die Basis aller praktischen Arbeiten und wurden durch eine intensive Literaturrecherche erarbeitet. Sie umfassen konstitutionelle Charakteristika von Cyanobakterien, die Analyse der speziellen Stoffwechselvorgänge und notwendigen Enzyme für die Produktion von molekularem Wasserstoff sowie eine umfassende Ermittlung wissenschaftlicher Erkenntnisse zu den Cyanobakterien *Synechocystis* sp. PCC 6803 und *Nostoc* sp. PCC 7120.

## 1.1 CYANOBAKTERIEN

Die ältesten Cyanobakterienfunde stammen aus Sedimentgesteinen und sind etwa 3,5 Mrd. Jahre alt. Damit gehören sie zu den ursprünglichsten Organismen auf unserem Blauen Planeten. Sie bevölkern beinahe alle Zonen der Erde und speziell angepasste Vertreter sind auch an Orten mit Extrembedingungen anzutreffen [Tamagnini *et al.*, 2002]. Sie sind zumeist photolithoautotroph und die einzigen Prokaryoten, welche die Eigenschaft einer pflanzenähnlichen oxygenen Photosynthese aufweisen [Fuchs & Schlegel, 2007]. Dadurch werden besonders die marinen Cyanobakterien zu einem der wichtigsten Bausteine der globalen CO<sub>2</sub>-Assimilation sowie O<sub>2</sub>-Produktion [Knoop *et al.*, 2010]. Indessen gibt es aber auch Arten, welche anoxygene Photosynthese ausüben können, d.h. sie haben die Möglichkeit manifestiert, Schwefelwasserstoff oder ähnliche Verbindungen an Stelle von Wasser als Elektronendonator zu verwerten [Tamagnini *et al.*, 2002]. Der Stamm der Cyanobakterien ist nach gegenwärtigem wissenschaftlichem Konsens in sieben Ordnungen eingeteilt, wobei noch ein großer Teil unklassifiziert geführt wird [URL-3]. Sie treten in der Umwelt in vielen verschiedenen Farbgebungen auf, was auf der heterogenen Ausprägung des photosynthetischen Pigments Chlorophyll a, den Carotinoiden sowie den Phycobiliproteinen basiert. Viele sind befähigt, elementaren Stickstoff aus der Atmosphäre biotisch zu fixieren, daraufhin Ammoniak zu synthetisieren und den Stickstoff somit besser für biochemische Reaktionen zur Disposition zu stellen. Dies erfolgt meist in Heterocysten, speziell differenzierten Stickstoff-fixierenden Zellen, welche von einem Großteil der filamentösen Cyanobakterien unter

Stickstoffmangel ausgebildet werden [Tamagnini *et al.*, 2002; Lindberg, 2003]. Ein ausgesprochen gut erforschtes Bakterium mit dieser Eigenschaft ist *Nostoc* sp. PCC 7120. Filamentöse Cyanobakterien charakterisieren ein spezielles Wachstumsmuster der Kolonien-bildenden Arten, die neben der Ausbildung in langen Filamenten auch bspw. als Aggregation zu Kugeln oder Matten existieren können. Außerdem gibt es weiterhin eine zweite Gruppe, die unizellulären, zu denen auch *Synechocystis* sp. PCC 6803 gehört. Bis vor kurzer Zeit war es allgemein anerkannt, dass alle Cyanobakterien über einen unvollständigen Zitronensäurezyklus verfügen, da ihnen das Enzym 2-Oxoglutarat-Dehydrogenase fehlt, sie somit nicht 2-Oxoglutarat zu Succinyl-CoA umwandeln können und deshalb auch kein Succinat synthetisiert werden kann. Diese Hypothese wurde von Zhang & Bryant im Jahr 2011 widerlegt. Sie wiesen nach, dass trotz der Abwesenheit von 2-Oxoglutarat-Dehydrogenase in fast allen Cyanobakterien eine Umwandlung von 2-Oxoglutarat zu Succinat über zwei weitere Enzyme möglich ist und der Zitronensäurezyklus somit als geschlossen interpretiert werden kann [Zhang & Bryant, 2011].

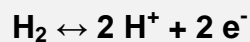
## 1.2 WASSERSTOFFPRODUKTION IN CYANOBAKTERIEN

Die Produktion von Wasserstoff durch Cyanobakterien kann durch zwei verschiedene Enzymkomplexe, die bidirektionale Hydrogenase sowie die Nitrogenase, katalysiert werden. Wasserstoff kann einmal den technischen Benchmark als Energieträger der Zukunft darstellen, wenn man es schafft, ihn ökonomisch zu erzeugen. Im Vergleich zu Erdöl hat es etwa den dreifachen Heizwert, flüssiger Wasserstoff hat nach der Kernenergie sogar die höchste massebezogene Energiedichte [Navarro *et al.*, 2009; Edwards *et al.*, 2007]. Cyanobakterien spielen dabei eine wirtschaftlich sowie wissenschaftlich bedeutsame Rolle neben diversen Algenarten [URL-7]. Die traditionelle Produktion von Wasserstoff basiert auf technischen Verfahren wie der Dampfreformierung oder das thermische Cracken von Erdgas sowie der Elektrolyse oder der thermochemischen Umsetzung von Wasser. Die Synthese durch Cyanobakterien oder Algen wird als Biophotolyse bezeichnet. Vergleichbare Ansätze sind weiterhin der fermentative Abbau chemischer Verbindungen sowie die Kombination

von fermentativen und photosynthetisch aktiven Bakterien als Hybrid-Systeme [Dasa & Veziroglu, 2001].

### 1.2.1 [NiFe]-HYDROGENASEN

Hydrogenasen katalysieren allgemein die Bildung sowie die Spaltung von molekularem Wasserstoff:

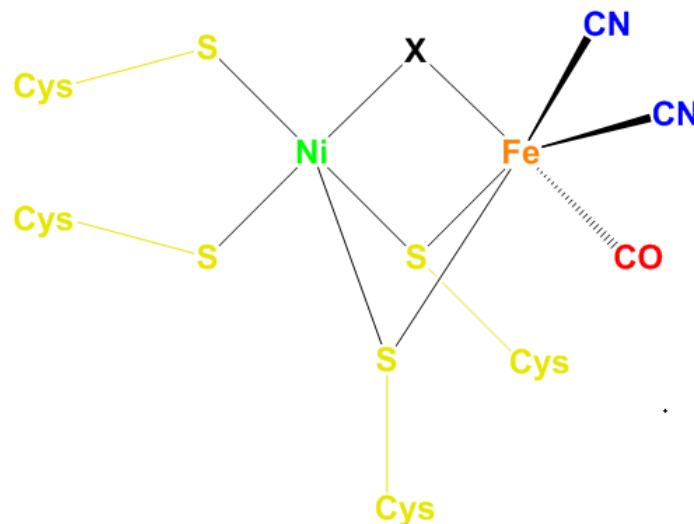


Gemäß der Zusammensetzung ihres aktiven Zentrums differenziert man zwischen [NiFe], [Fe] und metallfreien Hydrogenasen, wobei die [NiFe]- und [Fe]-Hydrogenasen über einen sehr ähnlichen chemischen und konstitutiven Aufbau verfügen. In Cyanobakterien sind nur [NiFe]-Hydrogenasen aktiv, welche ihre katalytische Wirksamkeit auf die zwei Metallatome, Nickel und Eisen, gründen und damit eines von vier Enzymen sind, welche Nickel im aktiven Zentrum koordinieren (Abbildung 1) [Tamagnini *et al.*, 2002, 2007; Lindberg, 2003; URL-7]. Sie können in vier Gruppen klassifiziert werden [Lindberg, 2003]:

<b>(Gruppe I)</b>	Membrangebundene respiratorische uptake [NiFe]-Hydrogenasen
<b>(Gruppe II)</b>	Zytoplasmatische heterodimere [NiFe]-Hydrogenasen
<b>(Gruppe III)</b>	Zytoplasmatische heterodimere reversible [NiFe]-Hydrogenasen
<b>(Gruppe IV)</b>	Membrangebundene H <sub>2</sub> -entwickelnde respiratorische [Ni-Fe]-Hydrogenasen

[NiFe]-Hydrogenasen sind immer als αβ-Heterodimer existent, wobei die α-Untereinheit das aktive Zentrum enthält und eine atomare Masse von etwa 60 kDa aufweist. Die β-Untereinheit besitzt nur etwa die halbe Masse und ist an der Übertragung von Elektronen beteiligt. Sie sind sehr empfindlich gegenüber der Anwesenheit von Sauerstoff und können bei zu hohen Konzentrationen Defizite in ihrer Funktionalität aufweisen [Tamagnini *et al.*, 2002, 2007; Lindberg, 2003; Hoffmann, 2008].





**Abbildung 1: Schematische Darstellung des aktiven Zentrums einer [NiFe]-Hydrogenase (nach Hoffmann, 2008)**

Darstellung einer [NiFe]-Hydrogenase im oxidierten inaktiven Zustand. Im aktiven Zentrum sind ein Nickel- sowie ein Eisenatom über zwei Schwefelatome von Cysteinresten miteinander verbunden. Das Nickelatom hat zudem Bindungen zu zwei weiteren Schwefelatomen, ist damit quadrat-pyramidal koordiniert, und das Eisenatom hat drei anorganische Liganden (zwei CN-Gruppen und ein CO-Molekül). Die Cyanidgruppen schützen wahrscheinlich das aktive Zentrum vor dem Einfluss von Sauerstoff. Das Hydridion ist als X visualisiert und wird als  $O^{2-}$ ,  $OH_2$ ,  $OH$  oder  $SO$  angenommen. In der reduzierten aktiven Form der Hydrogenase ist X gleich  $H^-$  [Hoffmann, 2008; URL-7].

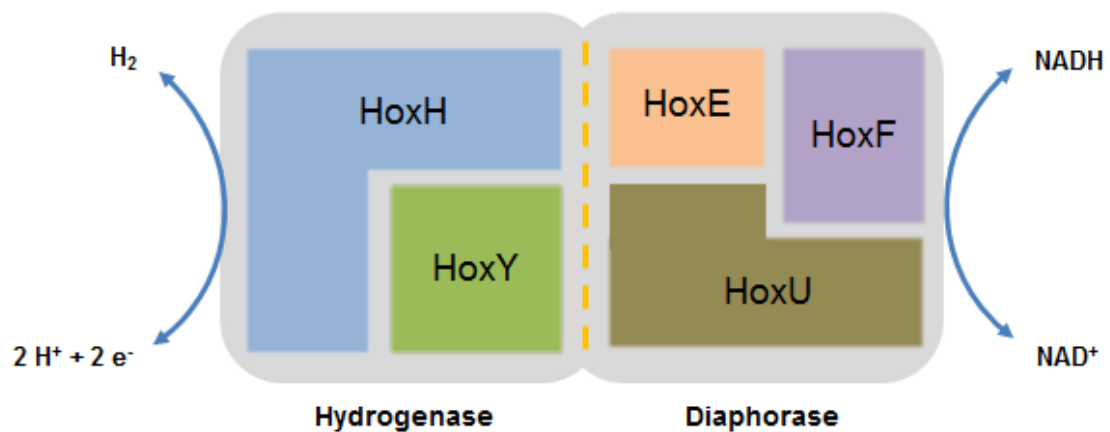
## BIDIREKTIONALE HYDROGENASE

Die bidirektionale Hydrogenase in Cyanobakterien ist der Gruppe III der [Ni-Fe]-Hydrogenasen zugehörig und katalysiert sowohl die Spaltung von Wasserstoff in Elektronen und Protonen als auch die Reaktion dieser zu molekularem Wasserstoff. Es ist ein Enzym, welches in Stickstoff-fixierenden Cyanobakterien, wie *Nostoc* sp. PCC 7120, und auch in jenen ohne diese Eigenschaft, wie *Synechocystis* sp. PCC 6803, vorkommt. Die physiologische Funktion ist trotz Jahren intensiver Forschung noch nicht vollends geklärt, jedoch werden verschiedene Annahmen dargelegt [Tamagnini *et al.*, 2002, 2007; Gutekunst *et al.*, 2005; Gutthann *et al.*, 2006; Agervald *et al.*, 2010]:

- Elektronen für die respiratorische Elektronentransportkette bereitstellen
- Elektronenventil, wenn Zellen fermentativ tätig sind
- Speicherung zusätzlicher Elektronen während der Adaption an eine größere Lichtintensität

- Beseitigung überschüssiger Reduktionsmittel unter fermentativen, anoxischen Bedingungen

Als Heterodimer sind zwei Bestandteile für eine funktionsfähige bidirektionale Hydrogenase essentiell, der Diaphorase-Teil, kodiert durch *hoxU* und *hoxF*, sowie der Hydrogenase-Teil. Dieser wird durch die zwei Gene *hoxY* und *hoxH* kodiert, wobei in den Cyanobakterien *Synechococcus* sp. PCC 6301, *Nostoc* sp. PCC 7120 sowie *Synechocystis* sp. PCC 6803 noch ein drittes Gen, *hoxE*, an der Expression beteiligt ist. Diese Form der bidirektionalen Hydrogenase ist in Abbildung 2 dargestellt [Tamagnini *et al.*, 2002, 2007; Lindberg, 2003].



**Abbildung 2: Die bidirektionale Hydrogenase als schematische Darstellung (nach Tamagnini *et al.*, 2007)**

Die Produktion von Wasserstoff wird im aktiven Zentrum, welche im Hydrogenase-Teil eingeschlossen ist, gewährleistet. Der Diaphorase-Teil wirkt entweder als Elektronendonator oder -akzeptor. Der Hydrogenase-Teil wird durch die Gene *hoxH* und *hoxY* kodiert, der Diaphorase-Teil durch *hoxU* sowie *hoxF*. In einigen Cyanobakterien, u.a. den Untersuchungsorganismen *Synechocystis* sp. PCC 6803 und *Nostoc* sp. PCC 7120, ist zusätzlich noch *hoxE* an der Expression des Diaphorase-Teils beteiligt.

Die zelluläre Lokalisierung der bidirektionalen Hydrogenase ist nicht vollständig geklärt. Es wird angenommen, dass sie entweder in den vegetativen Zellen und den Heterocysten ihre katalytische Wirkung entfaltet oder aber an der Zellmembran lokalisiert sind. Allgemein hin wird die bidirektionale Hydrogenase durch ihre hohe Wasserstoffaffinität, eine determinierte Thermotoleranz sowie ihre Sauerstoffsensitivität gekennzeichnet [Tamagnini *et al.*, 2002]. Außerdem

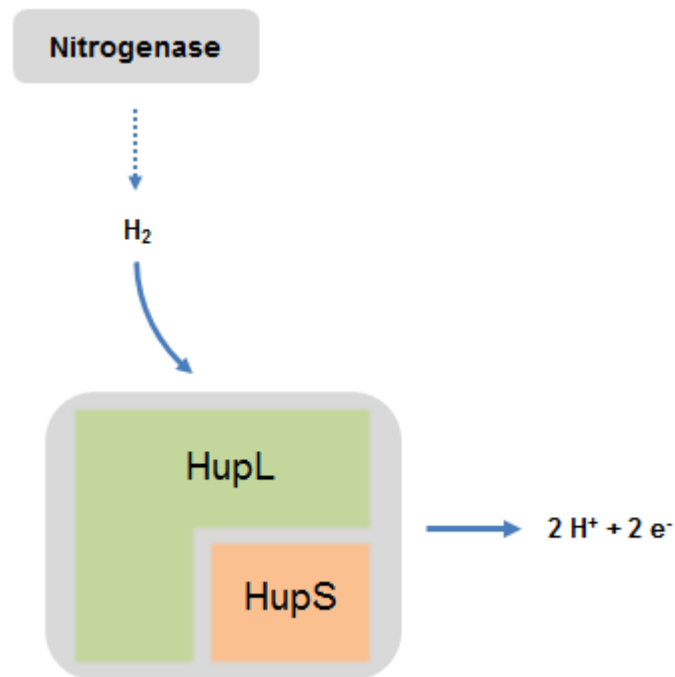
zeigt die bidirektionale Hydrogenase eine erhöhte Wirksamkeit unter anaeroben Bedingungen [Lindberg, 2003].

### UPTAKE HYDROGENASE

Die uptake Hydrogenase gehört zur Gruppe II der [NiFe]-Hydrogenase und wird gelegentlich auch als aufnehmende Hydrogenase bezeichnet [Lindberg, 2003]. Dieser Name kennzeichnet die Eigenschaft, Wasserstoff, welcher durch eine Nitrogenase gebildet wird, in Protonen sowie Elektronen zu spalten. Diese Funktionsweise schafft drei Vorteile für den Organismus:

- Bildung von ATP durch eine biochemische Knallgasreaktion
- Eliminierung von Sauerstoff und dadurch Schutz der Nitrogenase vor Inaktivierung durch ungebundenen Sauerstoff
- Bereitstellung von Reduktionsäquivalenten (Elektronen) für zelluläre Prozesse

Die große Untereinheit wird als HupL und die kleine als HupS bezeichnet, dargestellt in Abbildung 3. Außerdem werden diese gemeinsam transkribiert [Tammagnini *et al.*, 2002; Lindberg, 2003; Dutta *et al.*, 2005].



**Abbildung 3: Die uptake Hydrogenase als schematische Darstellung (nach Tamagnini et al., 2007)**

Die uptake Hydrogenase wird durch die Gene *hupS* und *hupL* kodiert und katalysiert die Spaltung von Wasserstoff zu Protonen und Elektronen. Der Wasserstoff stammt aus der Stickstofffixierung durch die Nitrogenase.

Da Nitrogenasen nur in Stickstoff-fixierenden Cyanobakterienstämmen vorkommen ist auch die uptake Hydrogenase fast ausnahmslos in solchen aufzufinden, aber auch dort nicht in allen. Die einzige Ausnahme bildet das nicht zur Stickstofffixierung befähigte Cyanobakterium *Synechococcus* sp. PCC 6301 [Tamagnini et al., 2002].

## AKZESSORISCHE PROTEINE

Enzyme, welche in dem aktiven Zentrum Nickel koordinieren, unterliegen einer intensiven Reifung, um biologisch wirksam zu sein. Die benötigten akzessorischen Proteine werden durch die *hyp*-Gene kodiert [Tamagnini et al., 2002, 2007; Hoffmann et al., 2006; Hoffmann, 2008]. Die bidirektionale Hydrogenase ist der Gruppe III-Hydrogenase von *E. coli* sequenziell sowie konformationell sehr ähnlich. Der Vorgang der Reifung ist bisher außerdem am intensivsten an den vier in *E. coli* vorhandenen [NiFe]-Hydrogenasen untersucht wurden, wes-

halb dieser im nachfolgenden kurz dargestellt wird. Der genaue Ablauf ist jedoch noch nicht bis ins Detail geklärt.

Der erste wesentliche Schritt ist die Übertragung der anorganischen Liganden CO und CN, welche an dem Komplex HypEF gebunden sind, an den Komplex HypCD und die Assemblierung des Eisenatoms welches an dem Komplex von HypC und HypD gebunden ist. Diese Anlagerung wird durch den Proteinkomplex aus HypE und HypF befähigt. Anschließend wird durch HypC das partiell assemblierte Metallzentrum an den Vorläufer von HoxH, dem sog. preHoxH, überführt und es bildet sich folglich ein Komplex aus HypC, preHoxH und dem teilweise assemblierten Metallzentrum. Der vorletzte Schritt beinhaltet den Einbau von Nickel ins resultierende vollständige Metallzentrum, welches durch den Komplex aus HypA1 und HypB1 inseriert wird. Anschließend wird über eine Endopeptidase, im Fall der bidirektionalen Hydrogenase von *Synechocystis* sp. PCC 6803 und *Nostoc* sp. PCC 7120 HoxW, der C-Terminus von preHoxH abgespalten und somit die fertige große Untereinheit des Hydrogenase-Teils der bidirektionalen Hydrogenase geformt [Hoffmann *et al.*, 2006; Hoffmann, 2008; Holmqvist, 2010]. Die Endopeptidase der großen Untereinheit der uptake Hydrogenase von *Nostoc* wird als HupW bezeichnet [Holmqvist, 2010]. Auch wenn ein Genom Gene für die erfolgreiche Expression der uptake und bidirektionalen Hydrogenase kodiert, wie es in *Nostoc* sp. PCC 7120 der Fall ist, ist stets in der Zelle nur ein Satz der *hyp*-Gene kodiert. Dies lässt auf eine Ko-Regulierung der *hyp*-Gene mit beiden Hydrogenasen schließen, welche allerdings experimentell noch nicht nachgewiesen wurde [Agervald *et al.*, 2008].

### 1.2.2 NITROGENASE

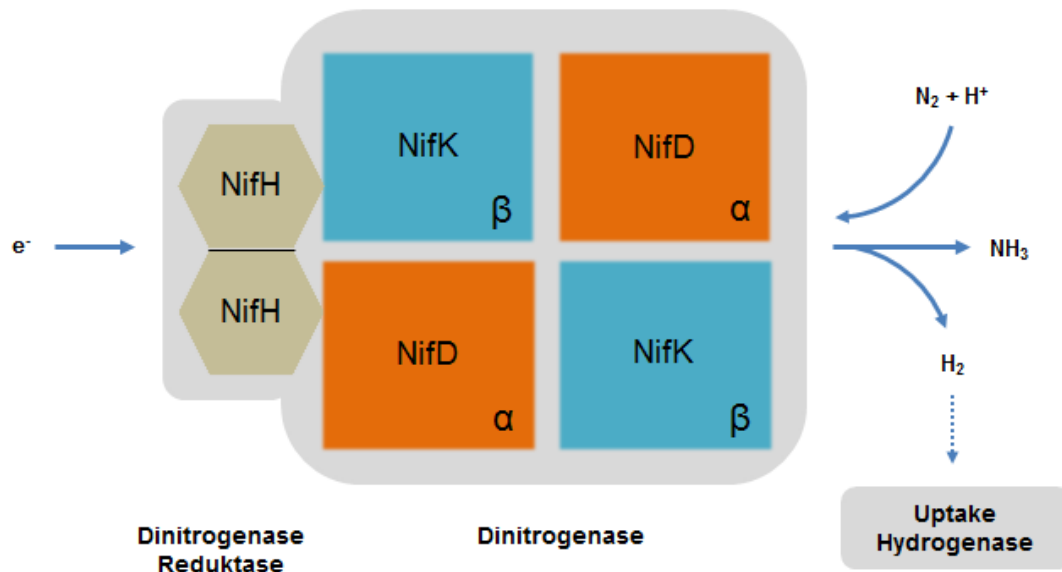
Die Nitrogenase ist ein Enzymkomplex aus einer Vielzahl von Untereinheiten und fungiert als eines der Schlüsselenzyme im Wasserstoffmetabolismus von Cyanobakterien. Der Komplex ist ausschließlich in Stickstoff-fixierenden Cyanobakterien aufzufinden und katalysiert dort die Reduktion von Stickstoff zu Ammonium unter ATP-Verbrauch. Als Nebenprodukt wird dabei stets mindestens ein Wasserstoffmolekül gebildet [Tamagnini *et al.*, 2002, 2007; Lindberg, 2003]. Außerdem kann ungebundener Stickstoff das Wachstum des Organis-

mus behindern [Tamagnini *et al.*, 2002]. Die allgemeine Reaktionsgleichung kann wie folgt formuliert werden [Tamagnini *et al.*, 2002; Hoffmann, 2008]:



Die Nitrogenase ist sehr sauerstoffempfindlich, weshalb die Stickstofffixierung am besten unter anaerobem Milieu stattfinden sollte. Viele filamentöse Cyanobakterien bilden deshalb Heterocysten, um eine räumlich abgeschlossene Stickstofffixierung unter anaeroben Konditionen durchzuführen. Eine weitere Methode, verbreitet bei Stickstoff-fixierenden Cyanobakterien die keine Heterocysten ausbilden können, ist die zeitliche Abgrenzung. Das bedeutet, der Organismus führt die Stickstofffixierung zu einem anderen Zeitpunkt wie die oxygene Photosynthese durch, weil der dort entstehende Sauerstoff sofort die Nitrogenase in ihrer Funktionsweise beeinträchtigen würde [Hoffmann, 2008; Tamagnini *et al.*, 2002].

In Cyanobakterien sind die Molybdän-Nitrogenasen vorzufinden, schematisch visualisiert in Abbildung 4. Sie bestehen aus einem Komplex von zwei Proteinen, der Dinitrogenase, auch als MoFe-Protein bezeichnet, und der Dinitrogenase-Reduktase (Fe-Protein). Die Dinitrogenase liegt stets als  $\alpha_2\beta_2$ -Heterotetramer vor und beinhaltet das aktive Zentrum sowie den in zwei Kopien vorliegenden FeMo-Cofaktor [Tamagnini *et al.*, 2002, 2007; Lindberg, 2003; Dutta *et al.*, 2005]. Um Elektronen aus der Dinitrogenase-Reduktase-Untereinheit zu empfangen, besitzt sie weiterhin zwei Kopien eines [8Fe-7S]-Clusters, welches auch als P-Cluster bezeichnet wird [Lindberg, 2003]. Die  $\alpha$ - und  $\beta$ -Untereinheiten der Dinitrogenase-Einheit werden durch *nifD* bzw. *nifK* kodiert, die Dinitrogenase-Reduktase wird als Homodimer ausgebildet und von dem Gen *nifH* kodiert. Die Dinitrogenase-Reduktase hat die Funktion, Elektronen von Elektron-Donoren, wie Ferredoxin oder Flavodoxin, zur Dinitrogenase-Einheit zu schleusen und somit die Umsetzung von Stickstoff zu gewährleisten [Tamagnini *et al.*, 2002, 2007; Lindberg, 2003; Dutta *et al.*, 2005; Kumar *et al.*, 2010]. Dafür ist das [4Fe-S]-Cluster zuständig [Lindberg, 2003].



**Abbildung 4: Die Nitrogenase als schematische Darstellung (nach Tamagnini *et al.*, 2007)**

Die Nitrogenase bildet sich aus den beiden Einheiten Dinitrogenase, kodiert als Heterodimer durch *nifD* und *nifK*, sowie Dinitrogenase-Reduktase, kodiert als Homodimer durch *nifH*. Die Nitrogenase befähigt die Umsetzung von Stickstoff zu Ammonium. Dabei entsteht als Nebenprodukt Wasserstoff, welcher in der Regel vollständig durch die uptake Hydrogenase umgesetzt wird.

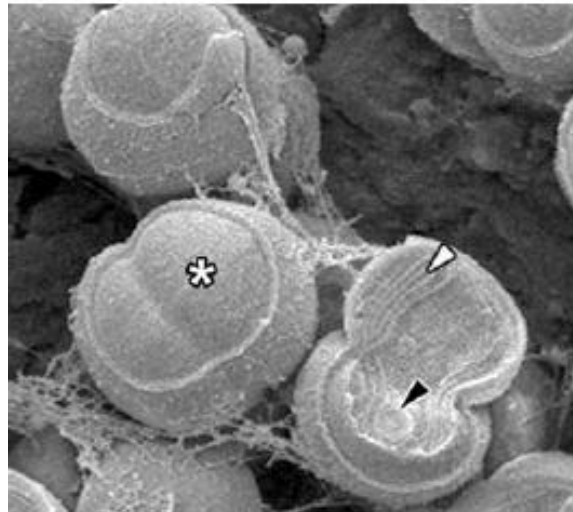
Es gibt Heterocysten-ausbildende Cyanobakterien, welche zusätzliche *nif*-Gene repräsentieren und somit auch unter sauerstofffreier Atmosphäre in den vegetativen Zellen eine Stickstofffixierung durchführen können [Kumar *et al.*, 2010]. Außerdem können zusätzliche *nif*-Gene für die Reifung der Nitrogenase kodiert sein [Masukawa *et al.*, 2010].

### 1.3 UNTERSUCHUNGSORGANISMEN

In der vorliegenden Arbeit wurden die zwei Cyanobakterien *Synechocystis* sp. PCC 6803 und *Nostoc* sp. PCC 7120 hinsichtlich ihrer involvierten Gene im Wasserstoffmetabolismus untersucht und die Erkenntnisse in metabolischen Modellen visualisiert.

### 1.3.1 *SYNECHOCYSTIS* SP. PCC 6803

Wenn im Bereich der Cyanobakterien von einem Modellorganismus die Rede ist, dann kann zum aktuellen wissenschaftlichen Momentum nur von *Synechocystis* sp. PCC 6803 gesprochen werden (Abbildung 5) [Montagud *et al.*, 2011; Anderson & McIntosh, 1991; Agervald *et al.*, 2010].



**Abbildung 5: Rasterelektronenmikroskopische Aufnahme von *Synechocystis* sp. PCC 6803 (nach van de Meene, 2005)**

*Zellverbund von Synechocystis sp. PCC 6803 als Kryo-Rasterelektronenmikroskopische Aufnahme. Der Stern deutet auf eine aufgebrochene Zelle zwischen der äußeren und der zytoplasmatischen Membran hin. Der weiße Pfeil indiziert das Thylakoid-Membransystem einer vollständig aufgebrochenen Zelle, der schwarze Pfeil einen zytoplasmatischen Einschluss.*

Bioinformatische Analysen und die Darstellung des Wasserstoffmetabolismus wurden in der vorliegenden Arbeit mit diesem einzelligen Organismus durchgeführt, weshalb dieser im Nachfolgenden genomisch näher charakterisiert wird und Grundlagen zu dem spezifischen Wasserstoffmetabolismus aufgegliedert werden.

#### GENOMISCHE DESKRIPTION

*Synechocystis* sp. PCC 6803 war der erste phototrophe Organismus, welcher vollständig sequenziert wurde [Sato *et al.*, 2007; Kaneko *et al.*, 1996]. Seit der Publikation dieser ersten grundlegenden wissenschaftlichen Errungenschaften im Jahr 1996 wurden viele neue Erkenntnisse gewonnen, sowohl auf Genom-



als auch auf Proteomebene. Deshalb ist es nicht verwunderlich, dass die anfangs eruierte Quantität an Genen für eine Proteinexpression um 149 auf nun 3317 angestiegen ist [Kaneko *et al.*, 1996; URL-8]. Die Bezeichnung der ORFs setzt sich aus drei Buchstaben sowie vier darauffolgenden Ziffern zusammen und resultiert aus der nachstehenden Typisierung [Kaneko *et al.*, 1995, 1996]:

1. **Artname** - im Fall von *Synechocystis* sp. PCC 6803 „s“
2. **Länge des ORF** - **l**  $\triangleq$  mehr als 100 Codons & **s**  $\triangleq$  weniger als 100 Codons
3. **Leserichtung** - **r**  $\triangleq$  direkter (codogener) Strang & **l**  $\triangleq$  Komplementärstrang (nicht-codogener Strang)
4. Folge von **vier Ziffern** im Bereich [0-9]

So heißt bspw. der ORF zu dem Protein Glucokinase, welches im ersten Schritt der Glykolyse der Umwandlung von  $\alpha$ -D-glucose zu  $\alpha$ -D-glucose 6-phosphate dient, in *Synechocystis* sp. PCC 6803: *slI0593*.

Cyanobakterien sind der einzige bekannte Bakterienstamm, welcher einen Tagesrhythmus bei der Expression bestimmter Gene aufweist. Dieser korrespondiert mit dem 24 h Tag-Nacht-Rhythmus und ist somit ein oszillierender Prozess. *Synechocystis* sp. PCC 6803 stellt durch diese Trennung den physiologischen Zustand der Zellen sicher, was bspw. Zellteilung oder Respiration beinhaltet. Auf molekularer Ebene beinhaltet dieser zirkadiane Rhythmus die positive und negative Rückkopplung bestimmter Gene, in Cyanobakterien der *kaiABC*-Gene. In einer umfassenden Studie wurden im Jahr 2005 in *Synechocystis* 54 Gene identifiziert, die dem zirkadianen Rhythmus unterliegen. Die meisten sind dabei an der Expression von Proteinen für den Kohlenhydrat- sowie Energiemetabolismus beteiligt [Kucho *et al.*, 2005].

## WASSERSTOFFMETABOLISMUS

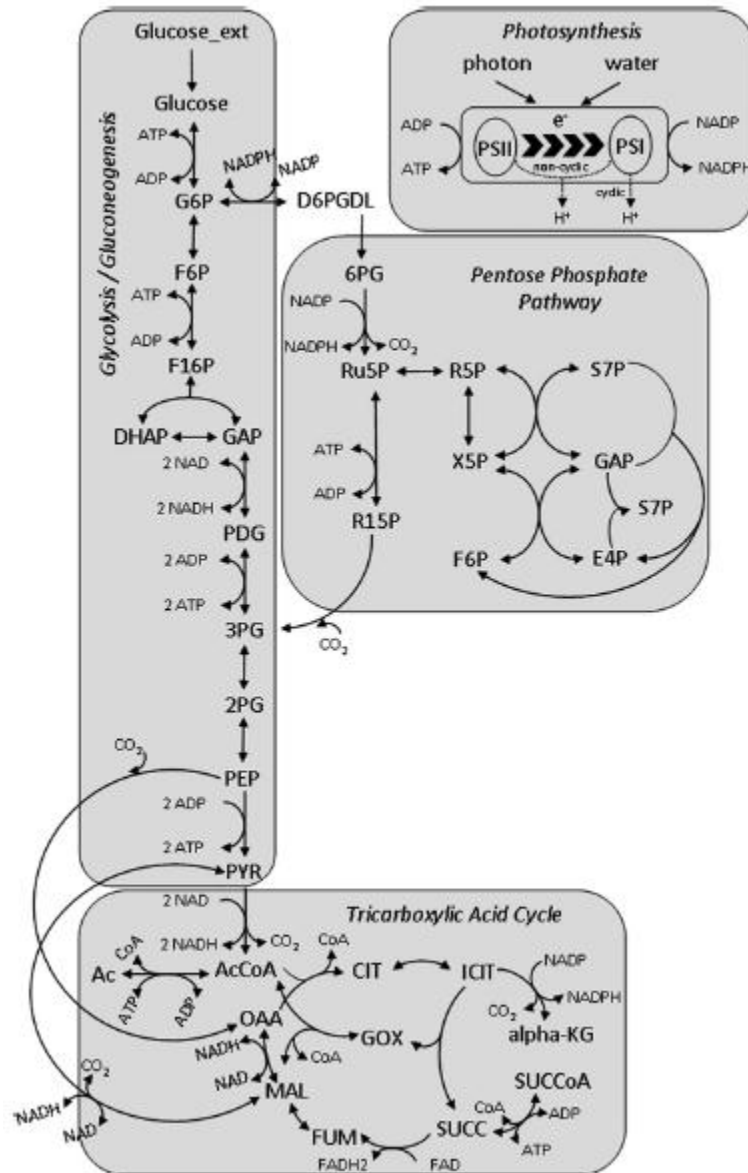
*Synechocystis* sp. PCC 6803 enthält als oxygener Organismus alle nötigen Bestandteile eines funktionsfähigen photosynthetischen Systems. Dies umfasst in besonderen Maße die Photosysteme I und II, wodurch die Oxidation von H<sub>2</sub>O zu O<sub>2</sub> realisiert wird und der zelluläre Energieträger ATP sowie weitere Reduk-

tionsäquivalente für metabolische Umsetzungen bereitgestellt werden [Burrows *et al.*, 2008; Montagud *et al.*, 2011]. Wasserstoff wird ausschließlich über eine bidirektionale Hydrogenase produziert, da *Synechocystis* sp. PCC 6803 als einzelluläres Cyanobakterium über keine Nitrogenase verfügt. Aus diesem Grund ist *Synechocystis* auch nicht befähigt, Stickstoff aus der Atmosphäre zu fixieren [Tamagnini *et al.*, 2002; Schütz *et al.*, 2004]. Unter stickstofffreier Atmosphäre kann es eine ähnliche Wasserstoffproduktivität wie Grünalgen oder filamentöse Cyanobakterien erreichen, sonst jedoch nicht. Dies resultiert aus der Anwesenheit von Sauerstoff zur Hemmung der Hydrogenase und bewirkt damit eine restringierte Produktion von Wasserstoff [Burrows *et al.*, 2008].

## DAS METABOLISCHE MODELL

Als Modellorganismus der Cyanobakterien wurden für *Synechocystis* sp. PCC 6803 bereits einige metabolische Analysen durchgeführt, mit dem Ziel des besseren Verständnis adaptierter Lebensweisen von Arten dieses *Phylums* [Navarro *et al.*, 2009; Montagud *et al.*, 2011; Knoop *et al.*, 2010; Yoshikawa *et al.*, 2011; Lopo *et al.*, 2012]. Es war von Beginn an evident, dass ohne Experimente im Umfang dieser Arbeit nur schwer essenzielle Reaktionen erfasst werden können. Deshalb diente für das in dieser Arbeit visualisierte Modell des Wasserstoffmetabolismus von *Synechocystis* sp. PCC 6803 das in der wissenschaftlichen Arbeit „Metabolic flux analysis of the hydrogen production potential in *Synechocystis* sp. PCC 6803“ validierte metabolische Konzept als Grundlage [Navarro *et al.*, 2009]. Dieses ist speziell mit der Absicht der umfangreichen Darlegung der Reaktionen, welche an dem Wasserstoffmetabolismus beteiligt sind, erforscht und verfasst worden. Eine Übersicht gibt Abbildung 6 wider [Navarro *et al.*, 2009]. Als Basis ihrer Analyse diente den beteiligten Wissenschaftlern die grundlegende metabolische Karte aus der *Kyoto Encyclopedia of Genes and Genomes*-Datenbank (KEGG) [URL-9]. Mittels der mathematischen Berechnungs- und Simulationsmethode *Flux Balance Analysis* (FBA) wurden alle individuellen Reaktionen und das Netzwerk quantifiziert. FBA nutzt die definierte Stöchiometrie der inter-, intra- und extrazellulären metabolischen Reaktionen als Input und liefert je nach simulierten Wachstumsbedingungen die übrigen metabolischen Informationen [Navarro *et al.*, 2009]. Das Modell wurde partiell

angepasst, da zudem teilweise progressive molekulare Erkenntnisse eingearbeitet wurden sowie die innere Konsistenz nicht an allen Stellen gemäß vergleichbarer Fachliteratur umgesetzt wurde.



**Abbildung 6: Metabolische Karte für *Synechocystis* sp. PCC 6803 nach Navarro *et al.* (2009)**

Die Darstellung des Wasserstoffmetabolismus basiert auf den dargestellten metabolischen Abläufen und wurde nur in wenigen Fällen an neue wissenschaftliche Erkenntnisse adaptiert. Primär betrachtete Stoffwechselwege sind die Glykolyse, der Zitronensäurezyklus sowie der Pentose-Phosphat-Stoffwechsel.

Die verschiedenen Stoffwechselwege für die Produktion molekularen Wasserstoffs sind stark an die photosynthetischen Vorgänge gekoppelt und funktionie-

ren nur im Zusammenspiel aller Komponenten. Die photosynthetische sowie respiratorische Elektronentransportkette verläuft an der Thylakoidmembran des Bakteriums. Die photosynthetische Elektronentransportkette ist stark mit dem zentralen Kohlenstoffkreislauf verkettet wobei durch metabolische Umwandlungsprozesse Elektronen freigesetzt werden, welche wiederum bei der respiratorischen Elektronentransportkette gebraucht werden. Stoffwechselwege, welche hauptsächlich Anteil an der Produktion von Wasserstoff haben, sind die biochemischen Umsetzungen in der Glykolyse, dem Zitronensäurezyklus sowie des Pentose-Phosphat-Stoffwechsels [Navarro *et al.*, 2009].

### BIDIREKTIONALE HYDROGENASE UND REGULATION

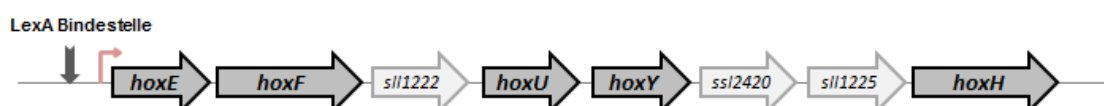
Wie bereits erwähnt, besitzt *Synechocystis* sp. PCC 6803 nur eine bidirektionale [NiFe]-Hydrogenase. Diese wird von fünf *hox*-Genen (*hoxEFUYH*) kodiert und gliedert sich in eine Diaphorase- (*hoxEFU*) sowie einen Hydrogenase-Untereinheit (*hoxYH*). Akzessorische Proteine für die Reifung werden in *Synechocystis* sp. PCC 6803 über die acht *hyp*-Gene *hypA1*, *hypB1*, *hypA2*, *hypB2*, *hypC*, *hypD*, *hypE* und *hypF* exprimiert, wobei *hypA2* und *hypB2* nur als Duplikate von *hypA1* und *hypB1* vorliegen und nach aktuellen Erkenntnissen nicht direkt an der Reifung beteiligt sind, eventuell aber Anteil an der Insertion der Eisenatome in das aktive Zentrum haben. Diese Gene sowie Proteine werden deshalb oft auch nur mit *hypA* und *hypB* sowie HypA und HypB bezeichnet. In dieser Arbeit wird jedoch wegen der Übersichtlichkeit mit *hypA1* und *hypB1* sowie den äquivalenten Proteinen HypA1 und HypB1 gearbeitet. Somit kann man konstatieren, dass in *Synechocystis* sp. PCC 6803 sechs Hyp-Proteine an dem Reifungsprozess der bidirektionalen Hydrogenase beteiligt sind. Die Gene sind anders als in vielen anderen Cyanobakterien im kompletten Genom verstreut [Tamagnini *et al.*, 2002, 2007; Hoffmann, 2008; Hoffmann *et al.*, 2006]. Tabelle 1 visualisiert die ORFs für die an der Hydrogenase beteiligten exprimierten sowie akzessorischen Gene.

**Tabelle 1: Auflistung der ORFs und zugehöriger Gen-Namen der bidirektionalen Hydrogenase-Untereinheiten sowie der akzessorischen Gene von *Synechocystis* sp. PCC 6803**

Die bidirektionale Hydrogenase von *Synechocystis* sp. PCC 6803 wird durch fünf Untereinheiten kodiert, welche den Hydrogenase-Teil (HoxYH) sowie den Diaphorase-Teil (HoxEFU) bilden. Die akzessorischen Hyp-Proteine sind für die Reifung der bidirektionalen Hydrogenase zuständig.

ORF	Gen-Name
<i>sll1226</i>	<i>hoxH</i>
<i>sll1224</i>	<i>hoxY</i>
<i>sll1223</i>	<i>hoxU</i>
<i>sll1221</i>	<i>hoxF</i>
<i>sll1220</i>	<i>hoxE</i>
<i>slr1496</i>	<i>hypD</i>
<i>ssl3580</i>	<i>hypC</i>
<i>ssl1462</i>	<i>hypE</i>
<i>sll0302</i>	<i>hypF</i>
<i>slr1675</i>	<i>hypA1</i>
<i>sll1432</i>	<i>hypB1</i>

Die Gene der bidirektionalen Hydrogenase werden als *hox*-Operon exprimiert und über die LexA-Bindestelle, welche upstream von *hoxE* gelegen ist, reguliert. Dargestellt ist diese genomische Ordnung in Abbildung 7 [Tamagnini *et al.*, 2002, 2007; Lindberg, 2003; Bothe *et al.*, 2010; Gutekunst, 2006; Hoffmann, 2008; Oliveira & Lindblad, 2007].



**Abbildung 7: *hox*-Operon von *Synechocystis* sp. PCC 6803 (nach Hoffmann, 2008)**

Die Expression der bidirektionalen Hydrogenase erfolgt in *Synechocystis* sp. PCC 6803 über fünf *hox*-Gene und wird über einen *lexA*-regulierten Promoter gesteuert.

LexA, kodiert durch den ORF *sll1626*, ist bisher der einzige in der Literatur bestätigte Transkriptionsfaktor für die Regulation der bidirektionalen Hydrogenase. In *E. coli* ist es ein Teil der SOS-Antwort zur DNA-Reparatur, in *Synechocystis* jedoch nicht direkt [Oliveira & Lindblad, 2005; Gutekunst *et al.*, 2005]. Es wurde gezeigt, dass LexA in *Synechocystis* sp. PCC 6803 nicht nur als Repressor, sondern auch als Aktivator agieren kann [Gutekunst, 2006; Oliveira & Lindblad, 2005]. LexA ist gemäß einigen Studien wahrscheinlich an der Thylakoidmembran koordiniert [Oliveira & Lindblad, 2007]. In einer Studie wurde weiterhin ein AbrB-ähnliches Protein im Genom von *Synechocystis* identifiziert (*sll0359*), welches als Regulator fungieren kann. Es besitzt eine ähnliche DNA-Bindungsdomäne wie das AbrB-Protein in *Bacillus subtilis* [Oliveira & Lindblad, 2007]. In diesem Organismus ist es bereits als ein vielseitiger Expressionsfaktor bekannt [Oliveira & Lindblad, 2007; Strauch, 1995]. In *Synechocystis* ist die genaue Funktion noch nicht identifiziert, jedoch weisen die Analysen darauf hin, dass es als Aktivator des *hox*-Genclusters wirken kann. Genauere Analysen müssten durchgeführt werden, um korrekte Aussagen für das Zusammenwirken von LexA und AbrB auf die Expression der bidirektionalen Hydrogenase zu treffen [Oliveira & Lindblad, 2007]. Außerdem wurde ein zweites AbrB-ähnliches Protein identifiziert (*sll0822*), welches nach ersten Erkenntnissen als negativer Regulator der Expression des *hox*-Operons agiert und damit eine verringerte Hydrogenaseaktivität impliziert [Dutheil *et al.*, 2012].

Verschiedene exogene Einflüsse sowie definierte Wachstumsbedingungen können eine gesteigerte Hydrogenaseaktivität hervorrufen (auszugsweise) [Baebprasert *et al.*, 2010; Hoffmann *et al.*, 2006]:

- Inkubation unter anaerober Atmosphäre für 24 h
- pH-Wert von 7,5 bei Raumtemperatur
- Präsenz von  $\beta$ -Mercaptoethanol
- gesteigerte Nickelkonzentration

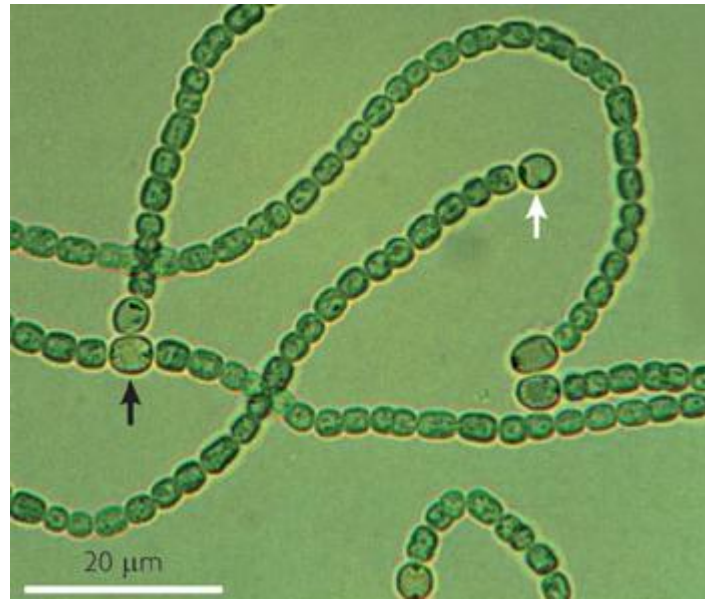
### 1.3.2 *NOSTOC* SP. PCC 7120

Neben *Synechocystis* sp. PCC 6803 gehört auch *Nostoc* sp. PCC 7120 zu einem der intensivsten erforschten Stämme der Cyanobakterien [Agervald *et*

*al.*, 2010; URL-28]. Auch zu diesem Organismus wurden intensive Analysen durchgeführt, um Charakteristika in Hinsicht auf die biologische Wasserstoffproduktion darzustellen, denn diese ist wesentlich umfangreicher als bei *Synechocystis*. Das Genom von *Nostoc*, ehemals auch als *Anabaena* bezeichnet, wurde im Jahr 2001 publiziert und weist mit etwa 6,4 Mio. chromosomalen Basenpaaren im Vergleich zu *Synechocystis* nahezu die doppelte Anzahl auf (Kaneko *et al.*, 2001; URL-8). In dieser wurde die Anzahl von 5368 identifizierten Protein-kodierenden Genen veröffentlicht, mittlerweile ist die potenzielle Anzahl im Organismus auf 5437 angestiegen [Kaneko *et al.*, 2001; URL-27]. Auch bei *Nostoc* setzt sich die Bezeichnung der kodierenden Gene wieder aus dem Artname (a), der Länge des ORF ( $l \triangleq$  mehr als 100 Codons,  $s \triangleq$  weniger als 100 Codons) sowie der Leserichtung ( $r \triangleq$  direkter Strang,  $l \triangleq$  Komplementärstrang) zusammen.

## VEGETATIVE ZELLEN VS. HETEROCYSTEN

*Nostoc* ist in der Lage Stickstoff aus der Umgebung zu fixieren, was den ATP-abhängigen Prozess der Reduzierung atmosphärischen Stickstoffs hin zu Ammonium beschreibt und durch die Nitrogenase katalysiert wird. Dies geschieht bei diesem Organismus durch die Ausbildung von Heterocysten, dargestellt in Abbildung 8. In Bezug auf diese Differentiation ist *Nostoc* am intensivsten erforscht. Das Protein NtcA, exprimiert durch das Gen *ntcA*, fungiert in allen Cyanobakterien als globaler Regulator des Stickstoffkreislaufs und wirkt dabei als Transkriptionsfaktor bei einer Vielzahl von involvierten Genen. Dabei hat es auch einen Einfluss auf die Ausbildung der Heterocysten. Experimente haben gezeigt, dass in *ntcA*-Mutanten die Differentiation blockiert wird [Olmedo *et al.*, 2006; Kumar *et al.*, 2010]. Ein weiterer elementarer Regulator bei der Differentiation der vegetativen Zellen zu Heterocysten ist das HetR. Auch Mutanten ohne dieses durch *hetR* exprimierte Protein sind nicht in der Lage Heterocysten auszubilden [Kumar *et al.*, 2010].



**Abbildung 8: Mikroskopische Aufnahme von *Nostoc* sp. PCC 7120 Filamenten [Florres & Herrero, 2010]**

Die Aufnahme zeigt Filamente von *Nostoc* sp. PCC 7120 unter Stickstoffmangel und die damit verbundene Ausbildung terminaler (weißer Pfeil) sowie interkalärer (schwarzer Pfeil) Heterocysten.

Zelluläre Besonderheiten von Heterocysten im Vergleich zu vegetativen Zellen sind unter anderem die Ausbildung von dickeren mehrschichtigen Zellwänden, die allgemein voluminösere Ausprägung sowie das häufige Vorhandensein von Cyanophycin-Körnchen an den Polen zu vegetativen Zellen. Die Heterocysten sind außerdem direkt von den vegetativen Zellen abhängig, da sie selbst kein Photosystem II besitzen und keine Kohlenstoffdioxid-Assimilation vollziehen können. Die vegetativen Zellen versorgen sie deshalb mit Kohlenhydraten und Reduktionsmittel. In Symbiose dazu übertragen die Heterocysten fixierten Stickstoff in Form von Aminosäuren. Durch diese Differentiation ist außerdem der Schutz der in den Heterocysten lokalisierten Hydrogenasen vor Sauerstoff gewährleistet [Tamagnini *et al.*, 2007].

#### INVOLVIERTE ENZYME IM WASSERSTOFFMETABOLISMUS

*Nostoc* sp. PCC 7120 repräsentiert die *nif*-Gene für die Expression der funktionstätigen Nitrogenase in verschiedenen Operons bzw. Genclustern. Die Dinitrogenase-Einheit wird durch die Gene *nifD* und *nifK* kodiert, die Dinitrogenase-Reduktase-Einheit durch *nifH* (Tabelle 2) [Tamagnini *et al.*, 2002; Kumar *et al.*,



2010]. Der Promoter des *nifHDK*-Operons ist nur in Heterocysten unter aeroben Konditionen aktiv [Flores & Herrero, 1994]. Weiterhin beinhaltet das primäre *nif*-Cluster ein *nif*-Operon upstream vom *nifHDK*-Operon, bezeichnet mit *nifB-fdxN-nifS-nifU*, sowie downstream die Gene *nifE-nifN-nifX-ORF-nifW-ORF-ORF-fdxH*. Außerdem repräsentiert das Genom von *Nostoc* ein weiteres separiertes Gencluster - *nifVZT* [Stricker *et al.*, 1997; Lindberg, 2003]. Einige dieser zusätzlichen *nif*-Gene sind als Cofaktoren für die erfolgreiche Expression der Nitrogenase zuständig, bei anderen ist die genaue Bedeutung noch nicht bekannt [Stricker *et al.*, 1997; Lindberg, 2003; URL-27]. Bis jetzt wurden noch keine Transkriptionsfaktoren in Cyanobakterien identifiziert, die für eine gesteuerte Expression der *nif*-Gene zuständig sind [Kumar *et al.*, 2010]. Der globale Stickstoffregulator NtcA kann allerdings als sequenzspezifisches Protein im Promoterbereich von *nifH* binden und könnte dadurch als Regulator wirken [Wei *et al.*, 1994].

**Tabelle 2: Auflistung der ORFs und zugehöriger Gen-Namen der Nitrogenase von *Nostoc* sp. PCC 7120**

Die Nitrogenase von *Nostoc* sp. PCC 7120 wird durch die beiden Untereinheiten Dinitrogenase und Dinitrogenase-Reduktase gebildet, exprimiert über die Gene *all1454* und *all1440* sowie *all1455*.

ORF	Gen-Name
<i>all1455</i>	<i>nifH</i>
<i>all1454</i>	<i>nifD</i>
<i>all1440</i>	<i>nifK</i>

*Nostoc* besitzt für den Wasserstoffmetabolismus weiterhin sowohl die bidirektionale als auch die uptake Hydrogenase [Tamagnini *et al.*, 2002, 2007; Agervald *et al.*, 2010]. Für deren Reifung sind wie in *Synechocystis* sp. PCC 6803 die akzessorischen Proteine zuständig, die durch die Gene *hypFCDEAB* kodiert werden, sowie die für die bidirektionale und uptake Hydrogenase individuell zuständigen Endopeptidasen HoxW und HupW, kodiert durch *hoxW* sowie *hupW*. Die Reifung der kleinen Untereinheit der uptake Hydrogenase ist noch nicht

geklärt [Holmqvist, 2010]. Eine Übersicht der Gen-Namen sowie der ORFs visualisiert Tabelle 3.

**Tabelle 3: Auflistung der ORFs und zugehöriger Gen-Namen der akzessorischen Gene von *Nostoc* sp. PCC 7120, die für die Reifung der bidirektionalen Hydrogenase entscheidend sind**

*Die Reifung der bidirektionalen Hydrogenase wird in *Nostoc* sp. PCC 7120 über sechs akzessorische Proteine gesteuert, die durch die Gene *hypFCDEAB* kodiert sind.*

ORF	Gen-Name
<b><i>alr0694</i></b>	<i>hypF</i>
<b><i>asr0695</i></b>	<i>hypC</i>
<b><i>alr0696</i></b>	<i>hypD</i>
<b><i>alr0698</i></b>	<i>hypE</i>
<b><i>alr0699</i></b>	<i>hypA</i>
<b><i>alr0700</i></b>	<i>hypB</i>
<b><i>all0770</i></b>	<i>hoxW</i>
<b><i>alr1423</i></b>	<i>hupW</i>

Die Gene *hypFCDEAB*, der zwischen *hypD* und *hypE* kodierte ORF *asr0697* sowie zwei downstream gelegene und fünf upstream kodierte ORFs scheinen als Operon exprimiert zu werden [Agervald *et al.*, 2010; Agervald *et al.*, 2008]. Studien mit dem Organismus *Rhizobium leguminosarum* bv. *Viciae* st. UPM 791 charakterisierten ein vier Gene umfassendes Cluster, *hupGHIJ*, welches an der Reifung der kleinen Untereinheit der uptake Hydrogenase beteiligt ist. Zwei Gene der fünf uptake gelegenen von *hypFCDEAB*, namentlich *alr0691* und *alr0692*, weisen konservierte Sequenzmotive auf, welche in den Genen *hupH* und *hupG* von *Rhizobium leguminosarum* vorzufinden sind. Diese Besonderheiten legen die Vermutung nahe, dass die upstream gelegenen Gene vom *hyp*-Cluster an der Reifung der kleinen Untereinheit der uptake Hydrogenase maßgeblich beteiligt sind [Agervald *et al.*, 2008]. Das Protein CalA reguliert als Expressionsfaktor HypC und wirkt dadurch auch direkt auf die erfolgreiche

Expression der bidirektionale Hydrogenase. Analysen zeigten, dass es bei Überexpression als Repressor wirkt. Nur unter Bedingungen der Stickstofffixierung werden in *Nostoc* beide Hydrogenasen exprimiert, die bidirektionale in den Heterocysten sowie in den vegetativen Zellen, die uptake Hydrogenase ausschließlich in den Heterocysten [Agervald *et al.*, 2010]. Die Gene der uptake Hydrogenase sind als Operon organisiert, die der bidirektionalen liegen in zwei benachbarten Clustern, *hoxEF* und *hoxUYH*, vor, werden jedoch nicht als Operon exprimiert. Die Transkriptionsstartpunkte für diese beiden etwa 8,8 kb voneinander entfernte Cluster liegen 66 sowie 57 bp upstream der Gene *alr0750* (upstream von *hoxE* gelegen) bzw. *hoxU* [Agervald *et al.*, 2010; Sjöholm *et al.*, 2007]. Übersichtlich sind die ORFs und Gen-Namen beider Hydrogenasen für den Organismus *Nostoc* sp. PCC 7120 in den Tabellen 4 und 5 dargestellt.

**Tabelle 4: Auflistung der ORFs und zugehöriger Gen-Namen der bidirektionalen Hydrogenase-Untereinheiten von *Nostoc* sp. PCC 7120**

Wie in *Synechocystis* sp. PCC 6803 wird auch in *Nostoc* sp. PCC 7120 die bidirektionale Hydrogenase durch die fünf Gene *hoxEFUYH* exprimiert.

ORF	Gen-Name
<b><i>alr0766</i></b>	<i>hoxH</i>
<b><i>alr0764</i></b>	<i>hoxY</i>
<b><i>alr0762</i></b>	<i>hoxU</i>
<b><i>alr0752</i></b>	<i>hoxF</i>
<b><i>alr0751</i></b>	<i>hoxE</i>

Analysen zeigten unter anderem, dass *hupL*-Mutanten, also eine *Nostoc*-Kultur, bei der die uptake Hydrogenase nicht exprimiert wird, bei kontinuierlicher Lichtbestrahlung bzw. optimalen Wachstumsbedingungen eine höhere Wasserstoffproduktion aufweisen als der Wildtyp sowie Kulturen, bei denen die bidirektionale Hydrogenase nicht vorhanden ist oder überhaupt keine Hydrogenase exprimiert wird [Marques *et al.*, 2011; Masukawa *et al.*, 2002].

**Tabelle 5: Auflistung der ORFs und zugehöriger Gen-Namen der uptake Hydrogenase-Untereinheiten von *Nostoc* sp. PCC 7120**

Die uptake Hydrogenase von *Nostoc* sp. PCC 7120 setzt sich aus den beiden Untereinheiten *hupS* und *hupL* zusammen, welche durch die Gene *all0688* bzw. *all0687* exprimiert werden.

ORF	Gen-Name
<b><i>all0688</i></b>	<i>hupS</i>
<b><i>all0687</i></b>	<i>hupL</i>

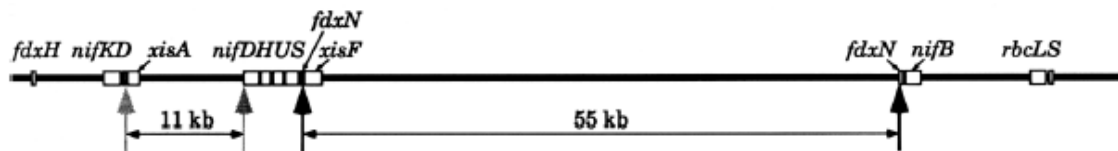
Auch in *Nostoc* gibt es das Regulatorprotein LexA. Studien haben ergeben, dass es mit den Promoterregionen der bidirektionalen Hydrogenaseeinheit interagiert. Ob es jedoch eine positive oder negative Expressionsrate hervorruft müssen Experimente noch klären [Sjöholm *et al.*, 2007]. Filamente von *Nostoc* bilden nur Heterocysten aus, wenn Stickstoff in limitierter Konzentration vorliegt. Bei dieser Differentiation wird etwa jede zehnte vegetative Zelle zu einem anaeroben Heterocysten. Diese auf die Fixierung von Stickstoff spezialisierten Zellen wachsen nicht mehr und unterliegen auch nicht mehr dem Zellteilungszyklus [Ramaswamy *et al.*, 1997; Carrasco *et al.*, 2005]. Die Filamente formen demnach ein eindimensionales Muster zwei unterschiedlicher Zelltypen aus, welches mit der einer Perlenkette vergleichbar ist [Kumar *et al.*, 2010; Carrasco *et al.*, 2005]. Dies verdeutlicht auch Abbildung 8 [Florres & Herrero, 2008]. Auf genomischer Basis werden während der späten Phase der Differentiation DNA-Elemente der drei Gene *fdxN*, *nifD* sowie *hupL* herausgeschnitten. Die dafür benötigten Rekombinasen werden von den *xis*-Genen exprimiert. Diese sind jeweils auf den DNA-Elementen kodiert, welche sie auch herausschneiden. Die Regulation ist noch nicht geklärt [Ramaswamy *et al.*, 1997; Carrasco *et al.*, 2005]. Tabelle 6 beinhaltet die Informationen dieser genomischen Reorganisationen.

**Tabelle 6: Auflistung der DNA-Elemente, die während der Heterocysten-Ausbildung verändert werden, die daran beteiligten Rekombinasen sowie die Beschreibung der genomischen Umordnung [Ramaswamy et al., 1997; Carrasco et al., 2005; Wei et al., 1994]**

Während der späten Phase der Differentiation von vegetativen Zellen hin zu Heterocysten werden die Gene *fdxN*, *nifD* sowie *hupL* reorganisiert. Für diese gezielte Rekombination sind, spezifisch für jedes DNA-Element, verschiedene Rekombinase-Proteine zuständig. Die Rekombination umfasst stets ein gezieltes Herausschneiden bestimmter DNA-Abschnitte in den Genen. Diese Abschnitte werden folglich so bezeichnet wie das Gen, auf dem sie sich befinden.

DNA-Element	Rekombinase	Art der Umordnung
<b><i>fdxN</i> (all1516)</b>	<i>xisF</i> (alr1459) <i>xisH</i> (alr1461) <i>xisI</i> (alr1462)	Herausschneiden eines 55 kb großen DNA-Fragments aus <i>fdxN</i> . Die Inaktivierung des <i>xisF</i> -Gens verhindert die Umordnung von <i>fdxN</i> . Experimente zeigten, dass auch das Herausschneiden eines 3,2 kb großen DNA-Fragments, in dem <i>xisHI</i> liegen, das Herausschneiden des <i>fdxN</i> -Elements verhindern kann, wodurch auf einen direkten Zusammenhang dieser Bereiche geschlossen werden kann.
<b><i>nifD</i> (all1454)</b>	<i>xisA</i> (alr1442)	Herausschneiden eines 11 kb großen DNA-Segments in <i>nifD</i> und damit verbundene Schaffung des funktionsfähigen <i>nifHDK</i> -Operons. Die Hemmung von <i>xisA</i> blockiert die Umordnung und es ist keine Stickstofffixierung möglich. In dem upstream Bereich von <i>xisA</i> kann außerdem NtcA binden und somit unter Umständen als ein Regulator für die geregelte Expression von <i>xisA</i> wirken.
<b><i>hupL</i> (all0687)</b>	<i>xisC</i> (alr0677)	Herausschneiden eines etwa 10,5 kb großen DNA-Fragments in <i>hupL</i> mittels der Rekombinase XisC, deren Gen an einem Ende des <i>hupL</i> -Elements kodiert ist. Wenn <i>xisC</i> nicht vorhanden ist, kommt es auch nicht zu dem Herausschneiden des <i>hupL</i> -Elements.

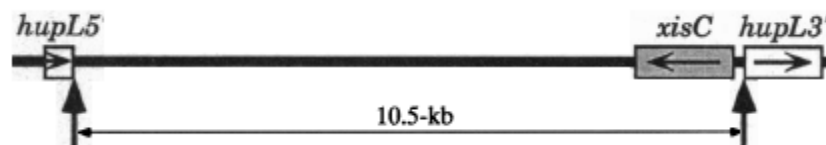
Die Darstellung der chromosomalen Abfolge der umstrukturierten DNA-Elemente *fdxN* und *nifD* visualisiert nachfolgend Abbildung 9. Dabei wird bei der Umordnung des *fdxN*-Elements nur die Möglichkeit mittels der *xisF*-Rekombinase und dem folgenden Herausschneiden des 55 kb großen DNA-Fragments in *fdxN* aufgeführt [Carrasco *et al.*, 1994].



**Abbildung 9: Schematische Darstellung der chromosomalen Abfolge im Bereich der umstrukturierten DNA-Elemente *nifD* und *fdxN* [Carrasco *et al.*, 1994]**

Darstellung der chromosomalen Abfolge der umstrukturierten DNA-Elemente *nifD* und *fdxN*. Für *fdxN* ist nur das Herausschneiden des 55 kb großen DNA-Elements mittels der Rekombinase *xisF* grafisch aufbereitet.

Des Weiteren visualisiert Abbildung 10 die genomische Beschaffenheit der Umordnung eines etwa 10,5 kb großen DNA-Elements in *hupL* durch die Rekombinase *xisC* [Carrasco *et al.*, 1995].



**Abbildung 10: Schematische Darstellung der chromosomalen Abfolge im Bereich des umstrukturierten DNA-Elements *hupL* durch die Rekombinase *xisC* (nach Carrasco *et al.*, 1995)**

Durch die Rekombinase *xisC* kommt es zu dem Herausschneiden eines etwa 10,5 kb großen DNA-Fragments in *hupL* während der Heterocysten-Differentiation.

## 2 MATERIAL UND METHODEN

Bioinformatische Software nimmt eine essenzielle Rolle in der modernen Biotechnologie, Bioverfahrenstechnik und der Genomanalyse ein. In der vorliegenden Arbeit sollten Reaktionen, welche am Wasserstoffmetabolismus von *Synechocystis* sp. PCC 6803 und *Nostoc* sp. PCC 7120 beteiligt sind, detektiert und mit einer geeigneten Software dargestellt werden. Im Folgenden werden in die Analysen einbezogene Datenbanken, die Visualisierungssoftware VANTED sowie der angewandte Visualisierungsstandard *Systems Biology Graphical Notation* (SBGN) näher erläutert.

### 2.1 SOFTWARE

#### 2.1.1 BIOINFORMATISCHE VISUALISIERUNGSSOFTWARE

Im Rahmen einer Kooperation mit der Bioinformatik-Abteilung des Leibniz-Instituts für Pflanzengenetik und Kulturpflanzenforschung in Gatersleben wurde während der Bearbeitungszeit für die Masterarbeit eine Studie über Visualisierungssoftware für die Darstellung metabolischer Netzwerke durchgeführt [URL-29]. Infolgedessen wurden 14 Softwaretools hinsichtlich ihrer

- Installationsprozedur;
- Anwendungsmöglichkeiten/Komplexität;
- Benutzerfreundlichkeit während des Visualisierungsprozesses

untersucht und verglichen. In Tabelle 7 werden alle getesteten Programme und deren Versionsnummern namentlich aufgeführt sowie das jeweilige Betriebssystem, auf dem sie installiert wurden.

**Tabelle 7: Auflistung der getesteten Softwaretools für die Visualisierung biologischer Netzwerke mit Angabe der Versionsnummer sowie dem Betriebssystem, auf dem die Software installiert und getestet wurde**

Programmname	Version	Betriebssystem
<b>BiNa</b> <sup>1</sup>	2.3.1	Windows
<b>BioTapestry</b> <sup>2</sup>	5.0.3	Windows
<b>BioUML</b> <sup>3</sup>	0.9.3	Linux
<b>Cell Designer</b> <sup>4</sup>	4.2	Linux
<b>Cytoscape</b> <sup>5</sup>	2.8.3	Windows
<b>MapMan</b> <sup>6</sup>	3.5.1	Windows
<b>Omix</b> <sup>7</sup>	1.5.19	Windows
<b>Ondex</b> <sup>8</sup>	ib08	Windows
<b>PathVisio</b> <sup>9</sup>	3.0	Windows
<b>Pathway Logic</b> <sup>10</sup>	3.4	Linux
<b>PathwayLab</b> <sup>11</sup>	1.3	Windows
<b>Pathway Projector</b> <sup>12</sup>	Beta	Web-Applikation
<b>VANTED</b> <sup>13</sup>	2.1.0	Windows
<b>VisANT</b> <sup>14</sup>	4.03	Windows

<sup>1</sup> [URL-31]

<sup>2</sup> [URL-32]

<sup>3</sup> [URL-33]

<sup>4</sup> [URL-34]

<sup>5</sup> [URL-35]

<sup>6</sup> [URL-36]

<sup>7</sup> [URL-37]

<sup>8</sup> [URL-38]

<sup>9</sup> [URL-39]

<sup>10</sup> [URL-40]

<sup>11</sup> [URL-41]

<sup>12</sup> [URL-42]

<sup>13</sup> [URL-1]

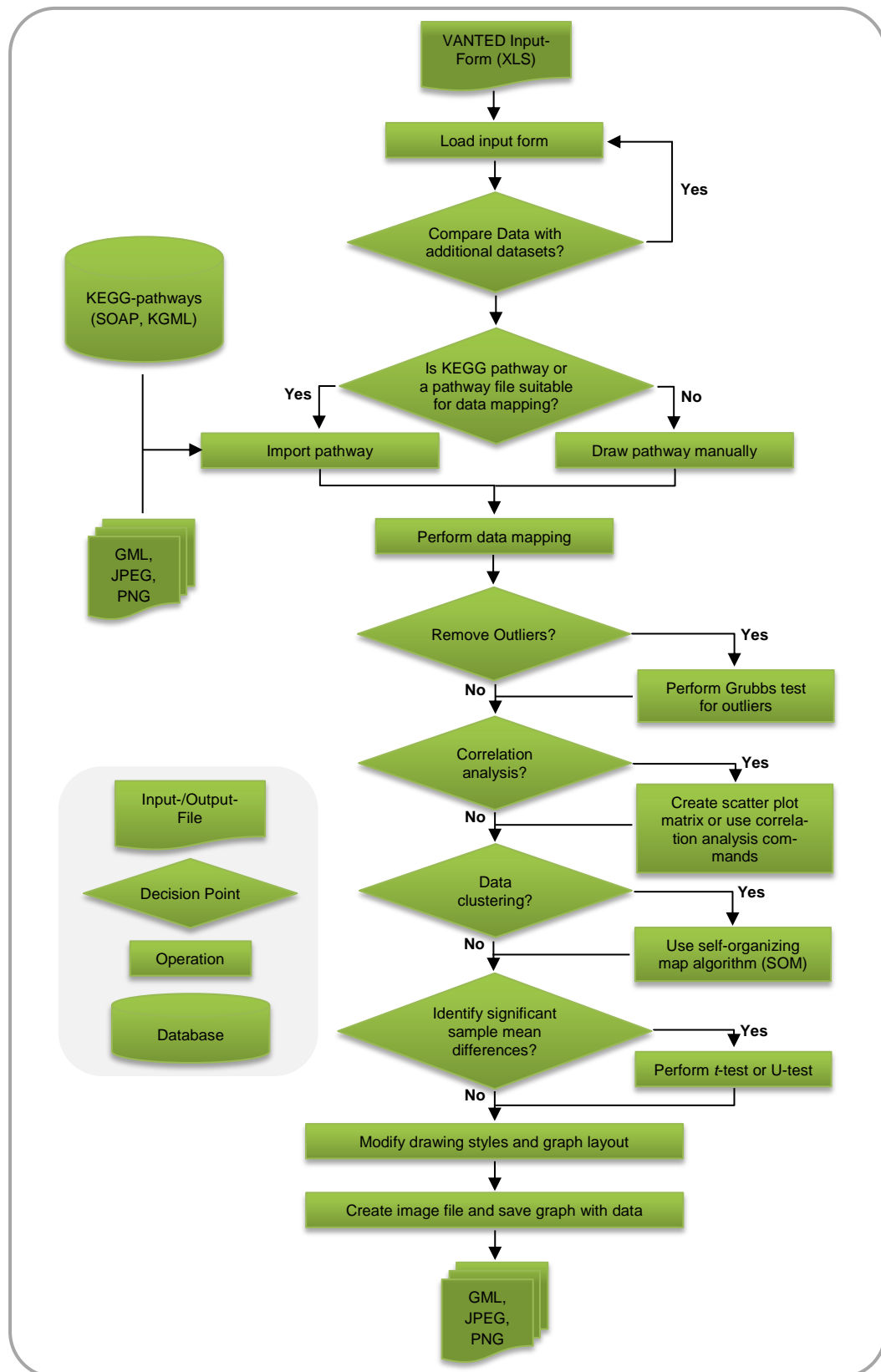
<sup>14</sup> [URL-43]



Die Programme wurden je nach Eignung und Verfügbarkeit auf dem Linux Betriebssystem Ubuntu, Version 12.04.1 LTS, oder Windows 7 Home Premium 64 Bit installiert und getestet. Das Windows-System befindet sich standardmäßig auf dem PC, das Linux Betriebssystem wurde unter dem Programm VirtualBox von Oracle, Version 4.1.20-80170, installiert [URL-30].

### 2.1.2 VANTED

Datenvisualisierung ist durch den Einsatz moderner Hochdurchsatzverfahren, wie Enzym-Assays oder Transcript-Profiling, eine der wichtigsten Aufgaben der modernen Bioinformatik. Außerdem müssen Programme zur automatischen Analyse der Messergebnisse entwickelt werden. Eines der bekanntesten auf diesem Forschungs- und Entwicklungsgebiet in den Pflanzenwissenschaften ist VANTED, prägnant für *Visualization and Analysis of Networks containing Experimental Data*. Es ermöglicht durch implementierte Prozeduren die mathematische Analyse und den Vergleich von mehreren importierten Datensätzen sowie das Mappen von Daten auf Netzwerke [Klukas *et al.*, 2006; Junker *et al.*, 2006; Klukas & Schreiber, 2010]. Die Möglichkeiten der Einbindung eigens implementierter Skripte in Java oder Ruby sowie die Integration von XML-Files der KEGG-Datenbank bestehen zusätzlich. Die Analyse und Darstellung biologischer Daten beginnt mit dem Import der Messdaten aus einem oder mehreren Microsoft Excel-Dokumenten. Diese Daten werden auf einen metabolischen Stoffwechselweg oder -kreislauf kartographiert. Dieser sollte entweder aus einer Datenbank wie KEGG stammen oder manuell erstellt worden sein. Anschließend können die Daten gemappt werden und es wird für jedes definierte Element ein Diagramm mit den entsprechenden Informationen angezeigt. Im Folgenden stehen mathematische Algorithmen zur Datenseparierung und Analyse zur Verfügung. Dabei werden im ersten Schritt mögliche statistische Ausreißer detektiert und entfernt. Anschließend können Datenkorrelationen erfasst, Separierungen realisiert und Signifikanztests durchgeführt werden. Der abschließende Schritt ist die Ordnung des Layouts des erstellten Netzwerks und das Exportieren in ein gewünschtes Ausgabeformat, wie GML, JPEG, PDF oder PPT [Junker *et al.*, 2006]. Der charakteristische Workflow einer Bearbeitung von Daten mittels VANTED ist in Abbildung 11 realisiert.



**Abbildung 11: Flussdiagramm - Arbeitsweise mit der Software VANTED (nach Junker et al., 2006)**

Nach dem Import von Datensätzen (Excel-Dokumente) werden die Informationen auf einen geeigneten Stoffwechselweg abgebildet. Anschließend beginnt die Analyse, bei der bspw. Da-

*tenausreißer entfernt werden können, verschiedene Korrelationen durchgeführt werden oder ein Clustering der Daten realisiert wird. Nach Bearbeitung des Netzwerkes und der dazugehörigen Daten kann das Netzwerk in verschiedene Formate exportiert werden.*

Die Möglichkeiten dieser Software wurden während der Bearbeitungszeit nicht im Entferntesten ausgeschöpft. Für vertiefende Arbeiten und umfangreichere Analysen sind experimentell ermittelte oder algorithmisch simulierte Daten nötig, welche leider nicht vorlagen. Aus diesem Grund ist es bei dem Erstellen manuell annotierter metabolischer Netzwerke geblieben. Dafür wurde zusätzlich der im Folgenden charakterisierte Notationsstandard *Systems Biology Graphical Notation* angewendet.

## 2.2 KONZEPTIONELLE VISUALISIERUNG

### **SYSTEMS BIOLOGY GRAPHICAL NOTATION**

Die Biologie in ihrer Vielseitigkeit zu erfassen ist eine sehr schwierige Aufgabe. Nicht weniger kompliziert ist der Ansporn, diese Diversität mit einer standardisierten Visualisierungsform grafisch darzustellen. Seit dem Jahr 2005 beschäftigt sich ein Team von Wissenschaftlern aus verschiedenen naturwissenschaftlichen und informatischen Abteilungen im Rahmen der Entwicklung der *Systems Biology Graphical Notation* (SBGN) mit dieser Thematik und nahm sich zur Aufgabe, einen Visualisierungsstandard in die Biologie einzuführen, wie er vergleichbar der Darstellung von elektronischen Schaltkreisen bereits seit Jahrzehnten zum Einsatz kommt [URL-2]. Das Projekt gründet sich auf eine Segmentierung in drei inhaltlich leicht divers konzeptualisierte Darstellungsformen, im nachfolgenden auch als Sprachen oder *language* bezeichnet. Hintergrund ist die adaptierte Deskription auf Grundlage des vorhandenen Wissens und Detailliertheit, d.h. wenn man nur den Einfluss biologischer Aktionen untereinander darstellen möchte, wählt man die *Activity flow language*, wenn Interaktionen zwischen Entitäten und deren Zusammenhänge visualisiert werden sollen die *Entity relationship language* und wenn die Absicht besteht ein umfassendes Netzwerk mit Veränderungen der Entitäten, chemischen Umwandlungen oder Lageänderungen von Entitäten zu beschreiben, die *Process*

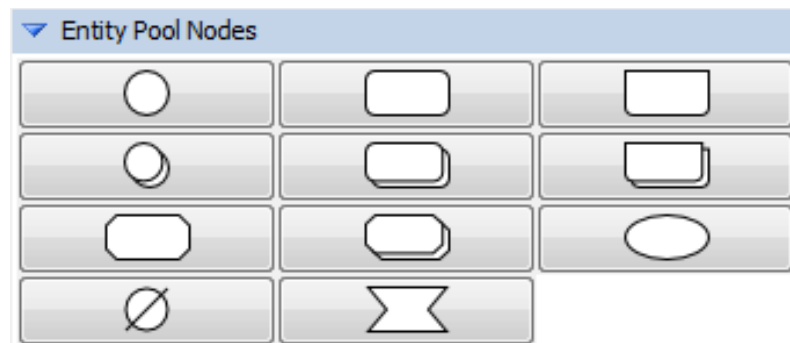
*description language* [Le Novère *et al.*, 2009; Beltrame *et al.*, 2011; Czauderna *et al.*, 2010; Moodie *et al.*, 2011; Mi *et al.*, 2009; Le Novère *et al.*, 2010]. Allen ist gemein, dass die Visualisierung in Form eines Graphen mit einer Auswahl verschiedener Kanten, Knoten und logischen Verknüpfungen realisiert wird [Le Novère *et al.*, 2009; Czauderna *et al.*, 2010]. Um standardisierte Netzwerke mit der SBGN-Notation zu erstellen braucht es selbstverständlich einen Editor. Dieser nennt sich SBGN-ED und ist als Add-On für VANTED verfügbar. Er ermöglicht der effektiven Nutzung der drei Sprachen, was die drei Hauptfunktionen:

- Erstellen und Editieren einer SBGN-Map
- Validieren einer SBGN-Map
- Übersetzung von Netzwerken in das SBGN-Format

umfasst. Der Import und die Übersetzung wird derzeit nur für Netzwerke von KEGG und MetaCrop gewährleistet [Czauderna *et al.*, 2010]. Einige Entwicklerteams von Softwaresystemen mit biologischem Hintergrund haben bereits die Intention von SBGN verinnerlicht und unterstützen in Ihren Tools diese Idee. Zusätzlich zu VANTED sind dies bspw. CellDesigner, BioUML oder die BioModels Database des EBI [Le Novère *et al.*, 2009].

Während des Forschungsmoduls wurde in SBGN als Sprache ausschließlich die *Process description language* benutzt. Der Grund dafür war die genaue Kenntnis über den Ablauf der biochemischen Reaktionen sowie der Hintergrund der Darstellung möglichst vieler Informationen. Die *Process description language* ermöglicht die Visualisierung eines Netzwerks mit sieben Hauptklassen an Bildzeichen:

- *Entity Pool Nodes*



**Abbildung 12: Übersicht über die *Entity Pool Nodes* der *Process description language***

*v. l. n. r.: Simple Chemical, Macromolecule, Nucleic Acid Feature, Multimer Simple Chemical, Multimer Macromolecule, Multimer Nucleic Acid Feature, Complex, Multimer Complex, Unspecified Entity, Source and Sink, Perturbing Agent*

- *Process Nodes*



**Abbildung 13: Übersicht über die *Process Nodes* der *Process description language***

*v. l. n. r.: Process, Omitted Process, Uncertain Process, Association, Dissociation, Phenotype*

- *Container Nodes*



**Abbildung 14: Übersicht über die *Container Nodes* der *Process description language***

*Im SBGN-Level 1, Version 1.3 ist nur der Container Node „Compartment“ verfügbar.*

- *Reference Nodes*



**Abbildung 15: Übersicht über die *Reference Nodes* der *Process description language***

v. l. n. r.: Submap, Tag right, Tag left, Tag down, Tag up

- *Connecting Arcs*



**Abbildung 16: Übersicht über die *Connecting Arcs* der *Process description language***

v. l. n. r.: Consumption, Production, Consumption/Production (Reversible), Modulation, Stimulation, Catalysis, Inhibition, Necessary Stimulation, Logic Arc, Equivalence Arc

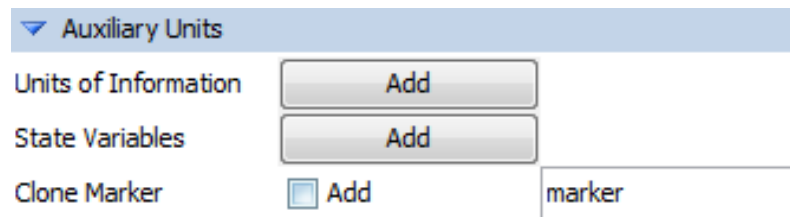
- *Logical Operators*



**Abbildung 17: Übersicht über die *Logical Operators* der *Process description language***

v. l. n. r.: And Operator, Or Operator, Not Operator

- *Auxiliary Units*



Auxiliary Units	
Units of Information	<input type="button" value="Add"/>
State Variables	<input type="button" value="Add"/>
Clone Marker	<input type="button" value="Add"/> <input type="text" value="marker"/>

**Abbildung 18: Übersicht über die *Auxiliary Units* der *Process description language***

Durch die Angabe von *Units of Information* oder *State Variables* können Entitäten definierte Eigenschaften verliehen werden. Ein Clone Marke kommt bei Entitäten dann zum Einsatz, wenn sie mehr als einmal in der Karte (bzw. *Compartment*) auftauchen. Dies dient der Übersichtlichkeit, besonders in umfangreichen Netzwerken.

Aktuell ist Level 1 mit der Version 1.3 verfügbar [Le Novère *et al.*, 2009; Moodie *et al.*, 2011].

## 2.3 DATENBANKEN

### 2.3.1 PROMBASE

Die regulierte Expression eines Genes oder gar eines gesamten Operons wird durch eine Nukleotidsequenz, welche meist upstream des exprimierten Gens bzw. Operons gelegen ist, gesteuert. Dieser DNA-Abschnitt wird Promoter genannt. Durch die ständig steigende Zahl an sequenzierten Genomen ist eine Progression des Wissens der regulatorischen Elemente von Bedeutsamkeit. In der Datenbank PromBase werden potentielle Promoterbereiche für 913 prokaryotische Genome verwaltet, die mittels eines Algorithmus berechnet und klassifiziert werden. Dieser stellt eine verbesserte Implementierung des PromPredict-Algorithmus dar [Kanhare & Bansal, 2005; Kanhare & Bansal, 2005; URL-22; URL-23]. Er verarbeitet für die Vorhersage die Stabilitätsunterschiede zwischen benachbarten upstream- und downstream-Bereichen in der Umgebung von experimentell erfassten Transkriptionsstartpunkten [Rangannan & Bansal, 2007, 2009]. Prognostizierte Promoter werden über alle 913 Genome mit einer Genauigkeit von 56 % korrekt vorhergesagt [Rangannan & Bansal, 2010]. Basierend auf einem Bewertungsschema für die Stabilität wird die Relia-

bilität für jeden vorhergesagten Promoter, basierend auf dem Unterschied zwischen seiner durchschnittlichen freien Energie und dem downstream-Bereich, in die fünf Klassen bzw. Level:

- *niedrig* - Level 1
- *durchschnittlich* - Level 2
- *hoch* - Level 3
- *sehr hoch* - Level 4
- *am höchsten* - Level 5

unterteilt [Rangannan & Bansal, 2010, 2011].

Im den in dieser Arbeit angefertigten metabolischen Modellen wurden mangels experimenteller Feststellungen aller Promoterbereiche für die meisten ORFs die Ergebnisse von PromBase eingebunden. Während der Literaturrecherche ist diese Datenbank positiv in Erscheinung getreten, außerdem gibt es nicht viele Programme auf diesem Forschungsbereich im Internet.

### 2.3.2 KEGG

Um einen ersten Überblick hinsichtlich der zellulären Abläufe, über vorliegende Stoffwechselwege und genomischen Deskription eines Organismus zu bekommen, ist die *Kyoto Encyclopedia of Genes and Genomes*-Datenbank seit Jahren eine der etabliertesten Anlaufstellen. Deren Entwicklung begann im Jahr 1995 im Zuge des japanischen „Human Genome Program“ [Ogata *et al.*, 1999]. Biochemische Daten aus wissenschaftlichen Studien und Experimenten sind in verschiedene Bereiche geclustert und visuell aufbereitet sowie teilweise zu anderen Datenbanken verlinkt [URL-16; Kanehisa *et al.*, 2004; Ogata *et al.*, 1999]. KEGG diente in der vorliegenden Arbeit dem Abgleich der biochemischen Reaktionen aus dem verarbeiteten metabolischen Modell von Navarro *et al.* (2009) sowie der Detektion exprimierter Gene und zugehöriger Enzyme in *Synechocystis* sp. PCC 6803 und *Nostoc* sp. PCC 7120.



### 2.3.3 BioCyc

Eine Vereinigung von speziellen Datenbanken zu einer umfassenden kann die Analyse von genomischen und metabolischen Daten signifikant verbessern, da man das Wissen zentralisiert abrufen kann. Eine solche Datenbank stellt BioCyc dar. Aktuell fließen Information aus 1962 Stoffwechsel- und Genomik-Datenbanken ein. Diese sind in drei Stufen unterteilt, welche sich auf die Validität stützen. Sechs Datenbanken gehören der Stufe 1 an. Diese sind mindestens ein Jahr durch Forscher literarisch verwaltet worden und werden mit wissenschaftlichem Progress intensiv erweitert. Hervorzuheben sind dabei besonders EcoCyc, HumanCyc und MetaCyc. Datenbanken der Stufe 2 und 3 werden ausschließlich nur mit verschiedenen Programmen generiert und bieten deshalb keine hervorragende Validität. Das Genom einer ausgewählten Spezies kann mittels einer visuellen Aufbereitung begutachtet werden und man kann sich je nach Interesse eine der vier Stufen:

- *sites*
- *genes*
- *operons*
- *genome*

anzeigen lassen [Karp *et al.*, 2005; URL-17]. Die Detektion von Operons wurde nicht mit BioCyc durchgeführt, da nicht ersichtlich ist, woher die entsprechenden Informationen stammen.

Die nicht-redundante MetaCyc-Datenbank enthält mehr als 1800 Stoffwechselwege und beteiligte Enzyme, welche aus über 30000 Publikationen abgeleitet wurden [Caspi *et al.*, 2011; URL-18]. Sie wurde besonders für die Evaluierung der Reaktionsrichtungen angewendet, da die KEGG-Datenbank darüber keine fundierten Informationen liefert.

### 2.3.4 SYNECHONET

Eine weitere Datenbank, welche auf die Vernetzung von mehreren Datenbanken vertraut, ist SynechoNET. Sie wurde speziell für die Detektion und Visuali-

sierung von Protein-Protein-Interaktionen, kurz PPI, in *Synechocystis* angelegt und greift auf gespeicherte Informationen von mehreren Portalen zu, unter anderem auf STRING, iPfam und CyanoBase [Finn *et al.*, 2005; Jensen *et al.*, 2008]. Es besteht die Möglichkeit der Darstellung mittels eines dynamisch implementierten Viewers, die Suche von Interaktionen mittels Genbezeichnung sowie die funktionell kategorisierte Suche durch die Einteilung der exprimierten Proteine in metabolische Prozesse. Es werden vorgefertigte Dokumente von detektierten Protein-Protein-Interaktionen aus den vier Datenbanken PSIMAP, iPfam, InterDom sowie STRING als Download bereitstellt [URL-21].

### 2.3.5 STRING

Protein-Protein-Interaktionen und deren Visualisierung sind stets von großer Bedeutung. Dadurch können Vorgänge wie Signaltransduktion und Transportprozesse innerhalb einer Zelle aufgeklärt werden. Eine entscheidende Aufgabe ist es deshalb, geeignete Experimente und Verfahren, computergestützte Algorithmen zur Vorhersage sowie verwaltende Datenbanken zu generieren. Experimentelle Methoden stellen bspw. das *Yeast Two-Hybrid System*, Phagendisplay und der Förster-Resonanzenergietransfer dar [Young, 1998; Sachdev *et al.*, 2003; Truong & Ikura, 2001]. Das *Search Tool for the Retrieval of Interacting Genes* (STRING) repräsentiert Informationen zu Protein-Protein-Interaktionen. In dieser Datenbank sind dabei sowohl experimentelle als auch algorithmisch vorhergesagte Daten hinterlegt. Interaktionen zwischen Proteinen werden über eine Bewertungsfunktion beurteilt und geben Aufschluss über die Genauigkeit der Vorhersage. Die Netzwerke sind dynamisch adaptierbar, wobei die Knoten Proteine repräsentieren und Kanten zugehörige Interaktion zwischen den Proteinen kennzeichnen. Diese Interaktionen können auf bis zu sieben unterschiedlichen Vorhersagemethoden basieren, weshalb zwischen den Knoten auch bis zu sieben unterschiedlich farbige Kanten existieren können. Jede Vorhersagemethode wird durch eine individuelle Farbe repräsentiert [Szkarczyk *et al.*, 2010; Jensen *et al.*, 2008; URL-15]. Die Datenbank beinhaltet mittlerweile Informationen zu über 1100 vollständig sequenzierten Organismen und liegt in der aktuellen Version 9.0 vor [Szkarczyk *et al.*, 2010; URL-15].

### 2.3.6 ODB

Die Operon DataBase (ODB) ist eine Datenbank mit dem Zweck der Archivierung und Vorhersage von Operons. Mittlerweile sind Informationen zu mehr als 1000 Genomen gelistet. In der Datenbank wird zwischen bekannten und putativen Operons unterschieden, wobei die bekannten der Literatur entnommen sind und die putativen mittels eines Algorithmus vorhergesagt werden. Dieser bezieht vier Freiheitsgrade ein:

1. intergenische Distanzen
2. funktionelle Beziehungen in biologischen Abläufen
3. Gen-Coexpression
4. Konservierung der Genfolge

Sie können neben der Default-Einstellung zudem selbst adaptiert werden, wodurch die Datenbank zusätzliche Ergebnisse liefern kann. Die Korrektheit der Vorhersagen müssten jedoch im Labor geprüft werden. Des Weiteren ist es möglich orthologe Gene zu detektieren, deren Daten von KEGG bezogen werden [Okuda *et al.*, 2006; Okuda & Yoshizawa, 2011; URL-11].

### 2.3.7 OPERONDB

Eine andere Datenbank für die Identifizierung von Operons in einem Organismus ist OperonDB. Gelistet sind in dieser Sammlung Operons von mittlerweile 1042 Organismen. Die Daten der Operon-Strukturen basieren ausschließlich auf algorithmisch bestimmten Ergebnissen, die sich auf die Konservierung von Genfolgen und deren Orientierung in mehr als zwei Spezies stützen. Der Output wird organismenspezifisch ausgewählt und umfasst Genpaare eines möglichen Operons. Diese Art der Datenrepräsentation wird aufgrund des verwendeten Berechnungsalgorithmus gewählt. Außerdem wird der kalkulierte Konfidenzwert sowie die Anzahl an homologen Genpaaren in anderen Genomen angegeben [Ermolaeva *et al.*, 2001; Perteau *et al.*, 2008; URL-19].

### 2.3.8 GENOMICS10

Ein weiterer Algorithmus zur Bestimmung von möglichen Operons stützt sich auf die Detektion von benachbarten Enzym-Clustern, die sukzessive chemische Reaktionen in metabolischen Stoffwechselwegen katalysieren. Biologischer Hintergrund ist dabei die Möglichkeit von Interaktionen zwischen Genen in Operons, entweder über die physikalische Interaktion ihrer Genprodukte oder durch vernetzt wirkende Reaktionen in Zellen, sodass sie wahrscheinlich an korrespondierenden Prozessen beteiligt sind. Algorithmisch wurde bei diesem Ansatz die Methodik der Graphentheorie verarbeitet [Zheng *et al.*, 2002]. Vorhersagen für 40 Genome werden mittels Vorhersage durch diesen Algorithmus in der Datenbank Genomics10 verwaltet [URL-12; URL-20].

## 2.4 KNOWLEDGEBASE

### **KNOWLEDGEBASE VON *MYCOPLASMA GENITALIUM* ALS VORLAGE FÜR CYANOBAKTERIEN**

Die *in silico* Modellierung metabolischer Umsätze in Organismen wird mit wachsender Kenntnis der biochemischen Umwandlungen stetig präziser, aber gleichzeitig auch umfangreicher. Um jedoch genomweite Berechnungen durchzuführen und nicht nur einzelne Stoffwechselabschnitte zu simulieren, ist auch ein umfassendes Berechnungsmodell von Vorteil. Lange waren besonders zwei Probleme existent, welche umfassende Modelle nicht möglich machten [Karr *et al.*, 2012]:

1. mangelndes Wissen über das Verhalten der Moleküle und ihrer Interaktionen
2. keine Verfügbarkeit einer algorithmischen Prozedur zur Berechnung der molekularen Komponenten und ihrer Interaktionen zu bestimmten zellulären Konditionen

Im Jahr 2012 wurde erstmals ein genomweites Modell des Urogenitalerregers *Mycoplasma genitalium*, gleichzeitig der Organismus mit dem kleinsten bekann-

ten Genom, etabliert, in welchem alle annotierten Funktionen der 525 Gene verarbeitet wurden [Fraser *et al.*, 1995; Karr *et al.*, 2012; URL-10]. Dies stellt ohne Frage einen enormen wissenschaftlichen Fortschritt dar, denn o.g. Probleme wurden erfolgreich überwunden und damit ist auch die Modellierung komplizierterer Organismen wie Cyanobakterien realistisch. Für das Modell für *Mycoplasma genitalium* wurde eine intensive Literaturrecherche durchgeführt, sodass Daten von über 900 Publikationen einbezogen wurden und sich diese in mehr als 1900 Parametern widerspiegeln [Karr *et al.*, 2012]. Diese Fülle an Informationen lässt auch vermuten, welche Arbeit bei der Etablierung einer *Knowledgebase* für einen komplexeren Organismus von Nöten ist, bspw. für *Synechocystis* sp. PCC 6803 mit derzeit bekannten 3317 Protein-kodierenden Genen [Kaneko *et al.*, 1996; URL-8]. Das Konzept des genomweiten Modells stützt sich auf die Unterteilung der zellinternen Prozesse in 28 miteinander agierende Module [Karr *et al.*, 2012]. In der Systembiologie, dem Wissenschaftszweig an der Schnittstelle zwischen u.a. Biologie, Mathematik, Physik und Informatik, wird dieser Modellierungsprozess als *bottom-up* Strategie bezeichnet [Kitano, 2002; Bruggeman & Westerhoff, 2006]. Da die molekularen-biochemischen Vorgänge der 28 Module nicht alle in gleicher Tiefgründigkeit erforscht und verstanden sind, werden für alle Module stets angepasste Modellierungsalgorithmen verwendet, welche manuell evaluierte Variablen und Werte repräsentieren. Die Standardmethode der biologischen Modellierung der letzten Jahren, die *Flux Balance Analysis*, wurde bspw. bei der Modellierung des Metabolismus genutzt [Karr *et al.*, 2012]. Leider ist mit derzeitigen wissenschaftlichen Kenntnissen und Methoden nicht zweifelsfrei beweisbar, dass alle modellierten Prozesse in der bestimmten Art und Weise in der Natur identisch ablaufen. Allein die Annäherung ist ein außerordentlicher wissenschaftlicher Erfolg und eröffnet weiterführende Möglichkeiten. Besonders die Identifizierung der Funktion von Genen und unbekannter genetischer Regulierungsmechanismen oder anderer molekularer Interaktionsprozesse kann richtungsweisende und spezialisierte Forschungsfelder eröffnen. Selbst die höchste Stufe, die Erstellung personalisierter Modelle für Menschen, rückt in greifbare Nähe. So könnte eine adaptierte Krankheitsbekämpfung durch individualisierte Medikamente und deren Konzentration sowie eine im hohen Maße gelenkte Genthera-

pie möglich werden. Leider ist die Schaffung einer umfassenden *Knowledge-base* eine im zeitlichen Umfang enorme Maßnahme, die sehr viel Expertenwissen in unterschiedlichen Forschungszweigen benötigt sowie ein bestimmtes technisches Know-how voraussetzt. Die Berechnungen für *Mycoplasma genitalium* wurden bspw. auf einem 128-Kern Linux-Cluster ausgeführt [Karr *et al.*, 2012].

## 2.5 STUDIE

### VISUALISIERUNGSSTANDARDS IN DER BIOLOGIE

„Finden Sie die Etablierung eines Visualisierungsstandards in der Biologie sinnvoll?“ Genau diese Frage sollte in der vorliegenden Arbeit ebenfalls erörtert werden, da im Umfang der Analyse vorhandener bioinformatischer Visualisierungssoftware mit einer Vielzahl unterschiedlicher Darstellungsformen gearbeitet wurde. Deren Fülle ist mittlerweile nur noch schwer überschaubar und viele Anbieter haben ihren eigenen Stil, Bildzeichen für definierte biologische Prozesse und Entitäten darzustellen. Die *Systems Biology Graphical Notation*, auszugsweise dargestellt unter Gliederungspunkt 2.2, soll deshalb einen Standard für die Zukunft setzen und die Prozesse der Biologie für alle Anwender in einer einheitlichen und verständlichen Sprache repräsentieren. Die Frage, wie die Forscher der Biowissenschaften zu diesem Sachverhalt stehen, wurde jedoch noch nie in den Fokus gestellt. Zu diesem Zweck wurde ein Fragebogen erstellt, welcher die Einstellung der Wissenschaftler zu einem Visualisierungsstandard und die Kenntnis von SBGN charakterisiert. Dazu wurde das Software-Paket „SoSci Survey“ eingesetzt [URL-25]. Dieses ist für den akademischen Nutzen kostenfrei einsetzbar und bietet eine übersichtliche und adäquat zu bedienende Benutzeroberfläche. Mit verhältnismäßig wenigen HTML-Kenntnissen lässt sich ein ansprechender und funktioneller Online-Fragebogen erstellen. Die Software beinhaltet außerdem die Möglichkeit des Exports des Umfrageergebnisses, so dass bspw. ein Import der Daten in Microsoft Excel möglich ist. Mit der Hilfe dieser Software wurde folglich ein Fragebogen erstellt, in dessen Kern vier prägnant formulierte Fragen zu der Thematik von Probanden beantwortet wer-

den sollten. Zu den Fragen 1 bis 3 gab es eine verbindliche Auswahl an Antwortmöglichkeiten, Frage 4 konnte individuell beantwortet werden. Die Fragen und ihre Antwortmöglichkeiten sind nachfolgend aufgeführt:

1. Finden Sie die Etablierung eines Visualisierungsstandards in der Biologie sinnvoll?  
ja / nein
2. Haben Sie bereits von dem Visualisierungsstandard SBGN gelesen oder sogar angewendet?  
nein / davon gelesen / bereits angewendet
3. Können Sie sich vorstellen, SBGN in Zukunft/weiterhin für Ihre Visualisierung anzuwenden?  
ja / nein / vielleicht
4. Stellen Sie explizite Ansprüche an einen Visualisierungsstandard? Gern können Sie nachfolgend einige Ihrer Wünsche äußern.

Der Fragebogen wurde außerdem noch in englischer Sprache ausformuliert und stand somit auch für anderssprachige Probanden zur Auswahl. Die potentiellen Kandidaten für die Studie wurden mittels Recherche von Professoren in Deutschland mit Lehrtätigkeit in bspw. Zellbiologie, Mikrobiologie, Biotechnologie, Bioinformatik, Biochemie oder Genetik sowie leitender Wissenschaftler in deutschen Instituten der Gesellschaften Max-Planck, Leibnitz und Helmholtz ermittelt. Insgesamt wurden dadurch über 300 Teilnehmer für die Studie identifiziert. Diesen wurde eine zu der Thematik ausformulierte E-Mail gesendet welche die Links zu dem deutsch- sowie englischsprachigen Fragebogen beinhaltete, sodass jeder Proband die freie Auswahl hatte. Des Weiteren wurde der Fragebogen über das soziale Netzwerk Facebook gezielt an geeignete Personen übermittelt [URL-26].

### 3 ERGEBNISSE UND DISKUSSION

Die erreichten Erkenntnisse sowie das Vorgehen im Umgang eingesetzter Datenbanken und Software werden im Folgenden geschildert.

#### 3.1 VISUALISIERUNG METABOLISCHER ABLÄUFE MIT VANTED SOWIE DER *SYSTEMS BIOLOGY GRAPHICAL NOTATION*

Während der Literaturanalyse wurden alle integrierten Details der in VANTED mittels SBGN konstruierten metabolischen Karten von *Synechocystis* sp. PCC 6803 sowie *Nostoc* sp. PCC 7120 evaluiert. Dabei kamen die VANTED-Version 2.1.0 und die SBGN *Process description language* Level 1, Version 1.3 zum Einsatz. Beide Modelle wurden auf unterschiedlicher Wissenslage erstellt, weshalb nachfolgend die individuellen Informationen aufgeführt werden.

##### 3.1.1 METABOLISCHE KARTE VON *SYNECHOCYSTIS* SP. PCC 6803

Die metabolische Karte von *Synechocystis* sp. PCC 6803 beinhaltet neben der Darstellung der Reifung der bidirektionalen Hydrogenase sowie der Synthese von ATP und NADP große Teile der Glykolyse, des Zitronensäurezyklus sowie des Pentose-Phosphat-Stoffwechsels. Die essentiellen Reaktionen stammen weithin aus dem Paper „Metabolic flux analysis of the hydrogen production potential in *Synechocystis* sp. PCC6803“, jedoch wurden Änderungen sowie Ergänzungen vorgenommen [Navarro *et al.*, 2009]. Änderungen betreffen teilweise die:

- Reaktionsrichtungen mittels der Datenbank BioCyc
- Modifikation der Stoffumsätze wie ATP auf Grundlage der Analysen mittels der Datenbanken BioCyc sowie KEGG
- Einbeziehung zusätzlicher metabolischer Zwischenschritte durch Evaluierung von Literatur sowie der KEGG-Datenbank



Ergänzend wurden

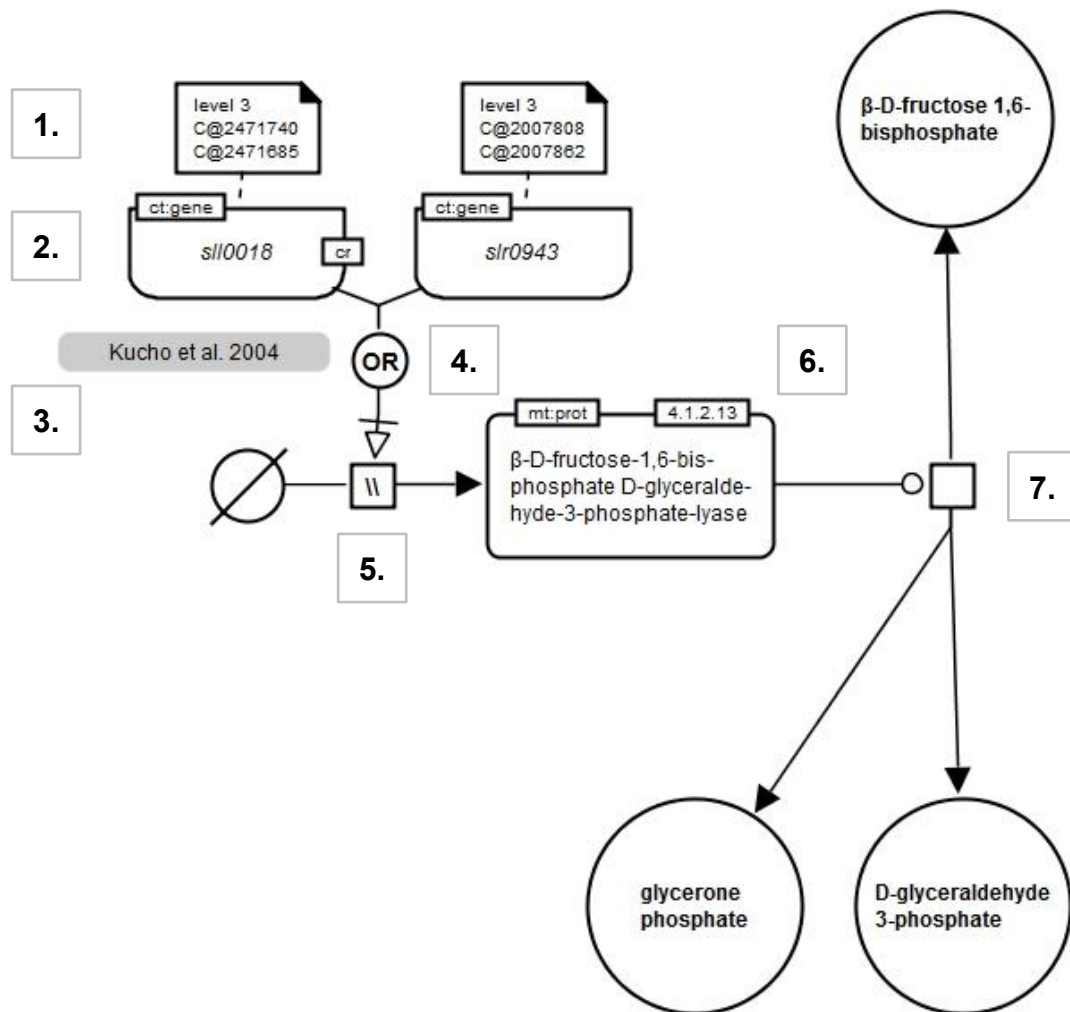
- die Reifung und der biochemische Stoffumsatz der bidirektionalen Hydrogenase, ermittelt durch die Literaturanalyse;
- putative sowie experimentell evaluierte Promoter-Bindestellen mittels PromBase sowie der berücksichtigten Literatur und
- Transkriptionsfaktoren, erfasst über die Literatur,

visualisiert. Zu begründen sind diese Modifikationen mit teilweise widersprüchlichen Ergebnissen in mehreren Datenbanken und anderen wissenschaftlichen Publikationen sowie der Hinzunahme weiterer enzymatischer Umsetzungen. In dem Modell wurden für die Expression der Gene die mit dem Programm PromBase ermittelten Promoter-Bindestellen integriert. Diesbezügliche Informationen sind im Anhang in Tabelle 13 abgelegt. Die visualisierten Informationen beinhalten die genomische Beschreibung des Starts und Endes der Bindestelle sowie das Level, welches die Reliabilität beschreibt (siehe 2.3.1). Alle Promoter-Bindestellen sind auf dem Chromosom lokalisiert. An einigen Stellen kann eine Proteinexpression durch mehrere Gene erfolgen. Dazu sind variierende ORFs mit dem booleschen Operator „OR“ logisch voneinander getrennt vermerkt und wenn vorhanden, die zugehörigen putativen Promoter-Bindestellen in jeweils einem *Reference Node* der *Entity relationship language* abgelegt. Diese Art von Bildzeichen ist derzeit noch nicht in der *Process description language* implementiert, soll aber schon bald auch in dieser Sprachen verfügbar sein und nach Rücksprache mit einigen der Entwickler ist es am sinnvollsten, Informationen dieser Art auf jene Weise zu visualisieren, auch wenn die vorliegende Karte dadurch prinzipiell zwar korrekt, jedoch nicht mehr valide ist. Die Expression der beteiligten Enzyme ist in einem kombinierten Schritt aus Transkription und Translation dargestellt. Dies ist besonders bei größeren Netzwerken sinnvoll, da trotz geringerer Details derselbe Informationsgehalt vorhanden ist und die Karte weniger unübersichtlich wird. Einzige Ausnahme bildet die Darstellung der drei durch die Literaturrecherche identifizierten Transkriptionsfaktoren *sll1626*, *sll0822* und *sll0359*, um ein unterscheidbares Kriterium festzulegen. In der Map sind wesentlich veränderte Stellen oder jene mit einer Besonderheit mit einem

gewöhnlichen *Reference Node* der *Process description language* versehen. Diese enthalten jeweils eine direkte Verlinkung auf die zugehörige Veröffentlichung in PubMed. Einige Gene des visualisierten Metabolismus von *Synechocystis* sp. PCC 6803 besitzen einen zirkadianen Rhythmus [Kucho *et al.*, 2005]. Im graphischen Modell sind diese mit dem Kürzel „cr“ als *Auxiliary Unit* im zugehörigen *Nucleic Acid Feature*, in dem der Genname hinterlegt ist, gekennzeichnet. Weiterhin ist dieser Datenknoten durch den zusätzlichen *Reference Node* „mt:gene“ eindeutig als Gen charakterisiert. Die exprimierten Enzyme sind als *Macromolecule Node* dargestellt und tragen als *Auxiliary Unit* den Vermerk „mt:prot“ in der linken oberen Ecke des Bildzeichens sowie die zugehörige EC-Nummer in der rechten oberen Ecke [URL-24]. In Abbildung 19 wird einmal beispielhaft der katalytische Prozess der biochemischen Umwandlung von  $\beta$ -D-fructose 1,6-bisphosphate zu D-glyceraldehyde 3-phosphate und glycerone phosphate durch das Enzym  $\beta$ -D-fructose-1,6-bisphosphate D-glyceraldehyde-3-phosphate-lyase dargestellt. Charakterisiert und beschrieben werden dabei zusammengefasst nachfolgende Informationen:

1. *Reference Node* der *Entity relationship language* mit Informationen zu jeweils einer möglichen Promoter-Bindestelle, dabei Angabe des Reliabilitätslevels sowie des Anfangs- und Endpunktes auf dem Chromosom (C@). Da sie in der *Process description language* noch nicht implementiert sind werden sie nach Validierung des Modells rot, also nicht valide, angezeigt, sind jedoch von der Darstellungsform korrekt.
2. *Nucleic Acid Feature* mit der Angabe möglicher Gene für die Expression eines Enzyms. Im Beispiel *sll0018* sowie *slr0943*. Das Gen *sll0018* wird in einem zirkadianen Rhythmus exprimiert, deshalb ist es mit der Kennzeichnung „cr“ versehen.
3. Verweis auf die Literaturquelle, in welcher nachgewiesen wurde, dass das Gen *sll0018* dem zirkadianen Rhythmus unterliegt.
4. Wenn wie in diesem Fall zwei verschiedene Gene zu der Expression des Enzyms führen können wird der logische Operator „OR“ eingesetzt.
5. Zusammenfassung der Umwandlungsprozesse Transkription sowie Translation.

6. *Macromolecule Node* für die Darstellung des exprimierten und katalytisch wirksamen Enzyms, hier  $\beta$ -D-fructose-1,6-bisphosphate D-glyceraldehyde-3-phosphate-lyase, sowie die Angabe der zugehörigen EC-Nummer [URL-24].
7. Visualisierung der biochemischen Umwandlung mit einem *Process Node*, in dem aufgeführten Szenario von  $\beta$ -D-fructose 1,6-bisphosphate zu D-glyceraldehyde 3-phosphate und glycerone phosphate.



**Abbildung 19:** Ausschnitt eines mit VANTED und des SBGN-Standards erstelltes metabolischen Netzwerk

1. Reference Node mit Information für eine putative Promoter-Bindestelle 2. Nucleic Acid Feature mit der Angabe möglicher Gene für die Expression eines Proteins, Kennzeichnung mit „mt:gene“ 3. Verweis auf die Literaturquelle, wenn die Komponente einer Besonderheit unterliegt. Im Bsp. unterliegt Gen sll0018 dem zirkadianen Rhythmus 4. logisches Oder, da das Protein durch zwei unabhängige Gene exprimiert werden kann 5. Prozess der Transkription/Translation 6. Macromolecule Node mit Angabe des Enzymnamen und dessen zugehöriger EC-Nummer [URL-24], Kennzeichnung mit „mt:prot“ 7. Process Node zur Kennzeichnung der katalytischen Reaktion, hier eine reversible Reaktion der Stoffe  $\beta$ -D-fructose 1,6-bisphosphate zu D-glyceraldehyde 3-phosphate und glycerone phosphate.

Die komplette metabolische Karte besitzt 515 Kanten und 595 Knoten, weshalb sie zu komplex und demzufolge zu groß ist, um sie im Umfang in dieser Masterarbeit grafisch abzubilden. Das Netzwerk wurde jedoch als Webseite exportiert und ist unter folgender Internetadresse abrufbar:

<http://www.hs-mittweida.de/wuenschi/SBGN/>

### 3.1.2 METABOLISCHE KARTE VON *NOSTOC* SP. PCC 7120

Da neben *Synechocystis* auch *Nostoc* sp. PCC 7120 in dem Fachgebiet der wirtschaftlichen Energieumwandlung mittels Cyanobakterien eine wissenschaftlich entscheidende Rolle spielt, wurde auch dessen Wasserstoffmetabolismus versucht darzustellen. Für *Nostoc* gibt es noch kein detailliertes metabolisches Modell des Wasserstoffmetabolismus, jedoch ist dieser Organismus in den dargestellten Stoffwechselwegen vergleichbar mit *Synechocystis*. An Hand dieses Wissens wurde die verfügbare Literatur nach Informationen durchsucht und vergleichende Analysen zu *Synechocystis* vollzogen. Für den Abgleich der enzymatischen Stoffumwandlungen wurden die jeweiligen annotierten Stoffwechselwege auf den Datenbanken KEGG und BioCyc verglichen [URL-16; URL-17]. Das Grundkonzept der grafischen Aufarbeitung der metabolischen Karte ist deshalb dasselbe wie bei *Synechocystis*. Zusätzlich sind jedoch die:

- Nitrogenase
- uptake Hydrogenase
- Reifung der uptake Hydrogenase
- Kompartiment des Heterocysten

visualisiert und mit verfügbaren Informationen zu Promoter-Bindestellen versehen. Diese wurden ebenfalls entweder direkt aus Veröffentlichungen übernommen oder aus der PromBase extrahiert und in einem *Reference node* vermerkt. Insgesamt umfasst diese Karte 715 Knoten und 614 Kanten. Diese sind verteilt auf die Kompartimente „vegetative Zelle“ sowie „Heterocyst“. Dabei verläuft die Stickstofffixierung durch die Nitrogenase sowie die gekoppelte Wasserstoff-

fumsetzung durch die uptake Hydrogenase im Heterocysten ab, alle weiteren Vorgänge sind in der vegetativen Zelle lokalisiert. Alle eingebrachten Gene und putative Promoter-Bindestellen sind wie für *Synechocystis* tabellarisch im Anhang aufgeführt (Tabelle 14). Die metabolische Karte für *Nostoc* sp. PCC 7120 ist durch die höhere Anzahl an Komponenten noch größer als die erstellte von *Synechocystis*. Aus diesem Grund ist auch diese nur auf nachfolgender Webseite abrufbar:

<http://www.hs-mittweida.de/wuenschi/SBGN/>

## 3.2 DETEKTION VON OPERONS

Für die Detektion vorhandener Operons in den betrachteten metabolischen Modellen von *Synechocystis* sp. PCC 6803 und *Nostoc* sp. PCC 7120 wurden die drei Datenbanken Genomics10, OperonDB und ODB eingesetzt. Die Analyse gestaltete sich aufgrund der genomischen Beschaffenheit der beiden Organismen als schwierig, da vorangegangene Analysen gezeigt haben, dass ihre Genome in verhältnismäßig wenigen Operons organisiert sind und auch häufig große Distanzen zwischen Genen eines Operons liegen [Moreno-Hagelsieb & Collado-Vides, 2002; Price *et al.*, 2005]. Ziel dieser Analyse war es festzustellen, in wieweit Transkriptionseinheiten in den metabolischen Modellen auftreten und folglich mittels Literatursuche mögliche Transkriptionsfaktoren festzustellen. Da die Literaturrecherche dieser Arbeit sich global hauptsächlich um Eigenschaften der biologischen Wasserstoffproduktion beschränkt, konnten natürlich nicht alle Komponenten der metabolischen Modelle ausführlich untersucht werden. Durch die Detektion der Operons war jedoch eine gerichtete Analyse möglich. Außerdem sollte durch dieses Vorgehen die Güte eingesetzter Operon-Datenbanken abgeschätzt werden.

Für *Synechocystis* sp. PCC 6803 kam es teilweise zu unterschiedlichen Ergebnissen. Genomics10 und OperonDB lieferten in drei Fällen vergleichbare Ergebnisse. ODB listet zwei putative Operons auf, jedoch handelt es sich dabei unverständlicherweise nur um einzelne Gene. Dieses Ergebnis ist wahrscheinlich auf den Algorithmus zurückzuführen. Trotz der Tatsache, dass diese Da-

tenbank auch Operons aus der Literatur vermerkt haben soll, ist von dieser Eigenschaft nicht auszugehen. Selbst vielfach in der Literatur beschriebene und erforschte Operons sind nicht in der Datenbank auffindbar. Validieren kann man uneindeutige Ergebnisse selbstverständlich nur durch eigens indizierte Experimente, sofern nicht bereits wissenschaftliche Publikationen darüber existieren und die Expression putativer Transkriptionseinheiten eindeutig nachgewiesen ist. Die Resultate der Analyse für *Synechocystis* sind in Tabelle 8 dargestellt. Nach der Analyse wurde mit den gewonnenen Ergebnissen zu den putativen Operons die Literatur nach Transkriptionsfaktoren untersucht. Für das putative Operon *slr0009-slr0011-slr0012* wurde herausgefunden, dass für die posttranslationale Regulierung der durch *slr0009* und *slr0012* exprimierten Untereinheiten der Ribulose-1,5-bisphosphat-Carboxylase wahrscheinlich hauptsächlich Anionen wie Orthophosphat verantwortlich sind [Marcus *et al.*, 2005]. Als Transkriptionsfaktor im klassischen Sinne ist dies jedoch nicht zu verstehen, weshalb diese Erkenntnis auch nicht in der metabolischen Karte von *Synechocystis* verarbeitet ist.

**Tabelle 8: Darstellung der identifizierten Operons von *Synechocystis* sp. PCC 6803 im betrachteten genomischen Ausschnitt mittels der drei Datenbanken Genomics10, OperonDB und ODB**

Die Datenbanken Genomics10 und OperonDB liefern in drei Fällen vergleichbare Ergebnisse der Vorhersage putativer Operons. Mittels ODB werden fälschlicherweise nur einzelne Gene vorhergesagt. Durch die Analyse der putativen Operons konnten neben den im Vorfeld durch die Literatur evaluierten Transkriptionsfaktoren, welche unmittelbar an der gesteuerten Expression von Genen für den Wasserstoffmetabolismus von *Synechocystis* beteiligt sind, keine weiteren Transkriptionsfaktoren identifiziert werden.

Genomics10	OperonDB	genomischer Abschnitt	ODB	
			Literatur	vorhergesagt
<i>sll1220, sll1221, sll1223</i>	<i>sll1224, sll1223, sll1221, sll1220</i>	1674350 - 1678565	keine Einträge	<i>sll1626</i>
<i>sll1322, sll1323, sll1324, sll1325, sll1326, sll1327, ssl2615</i>	<i>sll1322, sll1323, sll1324, sll1325, sll1326, sll1327, ssl2615</i>	171939 - 177523		<i>sll1432</i>
<i>slr0009, slr0012</i>	<i>slr009, slr0011, slr0012</i>	2478414 - 2480818		
<i>sll0018, sll0019</i>	<i>sll0402, sll0401</i>	2558820 - 2561305		
	<i>slr0193, slr0194</i>	2724648 - 2726033		

Für *Nostoc* konnte nur auf Informationen der beiden Datenbanken OperonDB sowie ODB zurückgegriffen werden, da Genomics10 diesen Organismus nicht vermerkt hat. In ODB waren drei vorhergesagte Operons aufgelistet, jedoch wie in *Synechocystis* keine aus bestehender Literatur, auch wenn bspw. das *nif*-Operon vielfach in der Literatur charakterisiert wird und theoretisch vermerkt sein müsste. Außerdem sind auch diesmal wieder nur einzelne Gen aufgeführt, welche jedoch nicht im Sinne eines Operons zu verstehen sind. OperonDB führt eine Vielzahl an Operons auf, wobei die meisten laut Literatur und Experimenten jedoch nicht exakt wie dargestellt exprimiert werden. Allerdings liefern die Daten eine gute Näherung. Beispielsweise wurden die Gene *alr0751* und *alr0752* als Operon gelistet, tatsächlich liegen sie jedoch als Operon mit nur einem weiteren Gen vor. Die Gene *alr0762*, *alr0764* sowie *alr0766* liegen als Operon mit zwei weiteren Genen vor [Sjöholm *et al.*, 2007]. Ein anderes Beispiel ist das Operon *atp1-atpI-atpH-atpG-atpF-atpD-atpA-atpC*, bestehend aus den ORFs *all0011-all0010-asl0009-all0008-all0007-all0006-all0005-all0004* [McCarn *et al.*, 1988]. In der Datenbank sind alle Gene exakt notiert, jedoch nicht das Gen *all0008*, namentlich mit *atpG* bezeichnet. Die einzige Erklärung dafür kann im Berechnungsalgorithmus der Daten für diese Datenbank liegen. Die Ergebnisse dieser Analyse sind in Tabelle 9 aufgearbeitet. Das putative Operon *all2563-all2564-all2566* wird in der Literatur als *xfp*-Operon bezeichnet und beinhaltet zusätzlich das Gen *all2567*. Für das Operon wurden putative Promoter-Bindestellen upstream von *all2567* sowie *all2563* detektiert und es wurde experimentell gezeigt, dass der Regulator NrrA (*all4312*) Einfluss auf die Expression dieser Gene hat [Ehira & Ohmori, 2011]. Aus diesem Grund ist er für die drei im Modell verarbeiteten Gene *all2563*, *all2564* sowie *all2566* mit eingearbeitet.

**Tabelle 9: Darstellung der identifizierten Operons von *Nostoc* sp. PCC 7120 im betrachteten genomischen Ausschnitt mittels der drei Datenbanken Genomics10, OperonDB und ODB**

Die Datenbank Genomics10 hat den Organismus *Nostoc* nicht gelistet. In der Datenbank ODB tauchen nur drei Treffer auf, welche nicht adäquat sein können. Verhältnismäßig zuverlässige Ergebnisse liefert die Datenbank OperonDB. In der konsekutiven Literatursuche wurde ein validierter Expressionsfaktor identifiziert. Dabei handelt es sich um das Protein NrrA, welches folglich für die drei Gene *all2563*, *all2564* und *all2566* als Transkriptionsfaktor in die metabolische Karte von *Nostoc* sp. PCC 7120 aufgenommen wird [Ehira & Ohmori, 2011].

Genomics10	OperonDB	genomischer Abschnitt	ODB	
			Literatur	vorhergesagt
keine Einträge	<i>all0004, all0005, all0006, all0007, asl0009, all0010, all0011</i>	3418 - 13521	keine Einträge	<i>alr4908</i>
	<i>alr0694, asr0695, alr0696, alr0698, alr0699, alr0700, alr0702</i>	800536 - 809010		<i>alr0700</i>
	<i>alr0751, alr0752</i>	872575 - 875014		<i>all3735</i>
	<i>alr0762, alr0764, alr0766</i>	883817 - 888250		
	<i>alr1041, alr1042</i>	1211965 - 1214507		
	<i>all1455, all1456, all1457, all1516</i>	1713396 - 1776518		
	<i>alr1524, alr1526</i>	1785970 - 1788279		
	<i>all2563, all2564, all2566</i>	3062538 - 3066504		
	<i>all5038, all5039, all5042, all5043, all5044</i>	6009574 - 6017468		

Jeder Biowissenschaftler wird wissen, dass Datenbanken für Analysen stets mit Bedacht gewählt werden müssen und die Daten bestenfalls noch einmal manuell evaluiert werden sollten, um die Etablierung von Fehlern in den eigenen Arbeiten zu vermeiden. Auch manuell annotierte Datenbanken können selbstverständlich Fehler enthalten, sind jedoch meist homogener in ihrer Daten- und Informationsstruktur [Bry & Kröger, 2003]. Auch verdeutlichen die Ergebnisse, dass man, wenn möglich, stets in mehreren Datenbanken nach Informationen suchen sollte. Das Ergebnis ist verblüffend und in seiner Deutlichkeit durchaus unerwartet. Durch die Identifikation eines weiteren Transkriptionsfaktors für die geregelte Expression mehrerer Gene, die effektiv im



Wasserstoffmetabolismus von *Nostoc* sp. PCC 7120 ist die Analyse als Erfolg zu bezeichnen.

### 3.3 PROTEIN-PROTEIN-INTERAKTIONEN

Es wurde versucht, vorhandene Protein-Protein-Interaktionen innerhalb der untersuchten metabolischen Netzwerke von *Nostoc* und *Synechocystis* unter Anwendung der *Systems Biology Graphical Notation* zu visualisieren. Dafür wurde anfangs für *Synechocystis* die Datenbank SynechoNET auf ihren Informationsgehalt hin überprüft. Der Import bereitgestellter Dokumente von Protein-Protein-Interaktionen in Cytoscape funktionierte einwandfrei, jedoch ist es nicht gelungen, diese sinnvoll aufzuarbeiten und darzustellen. Problematisch ist außerdem, dass keine Metainformationen zu den Interaktionen bezogen werden können und man somit nicht weiß, woher die Daten stammen. Um dieses Problem zu umgehen wurde folglich die STRING-Datenbank in die Analysen integriert [Jensen *et al.*, 2008; URL-15]. Diese Datenbank ermöglicht durch ihre API auch das Abrufen von Daten ohne die grafische Weboberfläche. Dafür muss ein *Uniform Resource Identifier*, kurz URI, generiert werden, welcher alle Angaben zum erzeugten Output beinhaltet. Dieser setzt sich folgendermaßen zusammen:

**`http://[database]/[access]/[format]/[request]?[parameter]=[value]`**

Für alle Bezeichner **[identifizier]** können gemäß einer Liste verschiedene Werte eingesetzt werden, sodass unterschiedliche Ansprüche bedient werden können. In den erstellten metabolischen Karten sind alle URI für jegliche ORFs derartig wie nachfolgend für den ORF *slr0752* aufgebaut:

**`http://string-db.org/api/image/network?identifiers=slr0752`**

Dieser URI beinhaltet Bezeichner mit nachfolgenden Angaben:

**1. [database]  $\triangleq$  string-db.org**

Zugriff auf die „string-db.org“ vornehmen. Neben dieser gibt es noch zwei weitere, namentlich „string.embl.de“ und „stitch.embl.de“.

**2. [access]  $\triangleq$  api**

Ansteuern der API von der ausgewählten Datenbank.

**3. [format]  $\triangleq$  image**

Angabe des Rückgabeformats, es soll ein Bild zurückgegeben werden.

**4. [request]  $\triangleq$  network**

Die Anforderung soll ein Netzwerk darstellen.

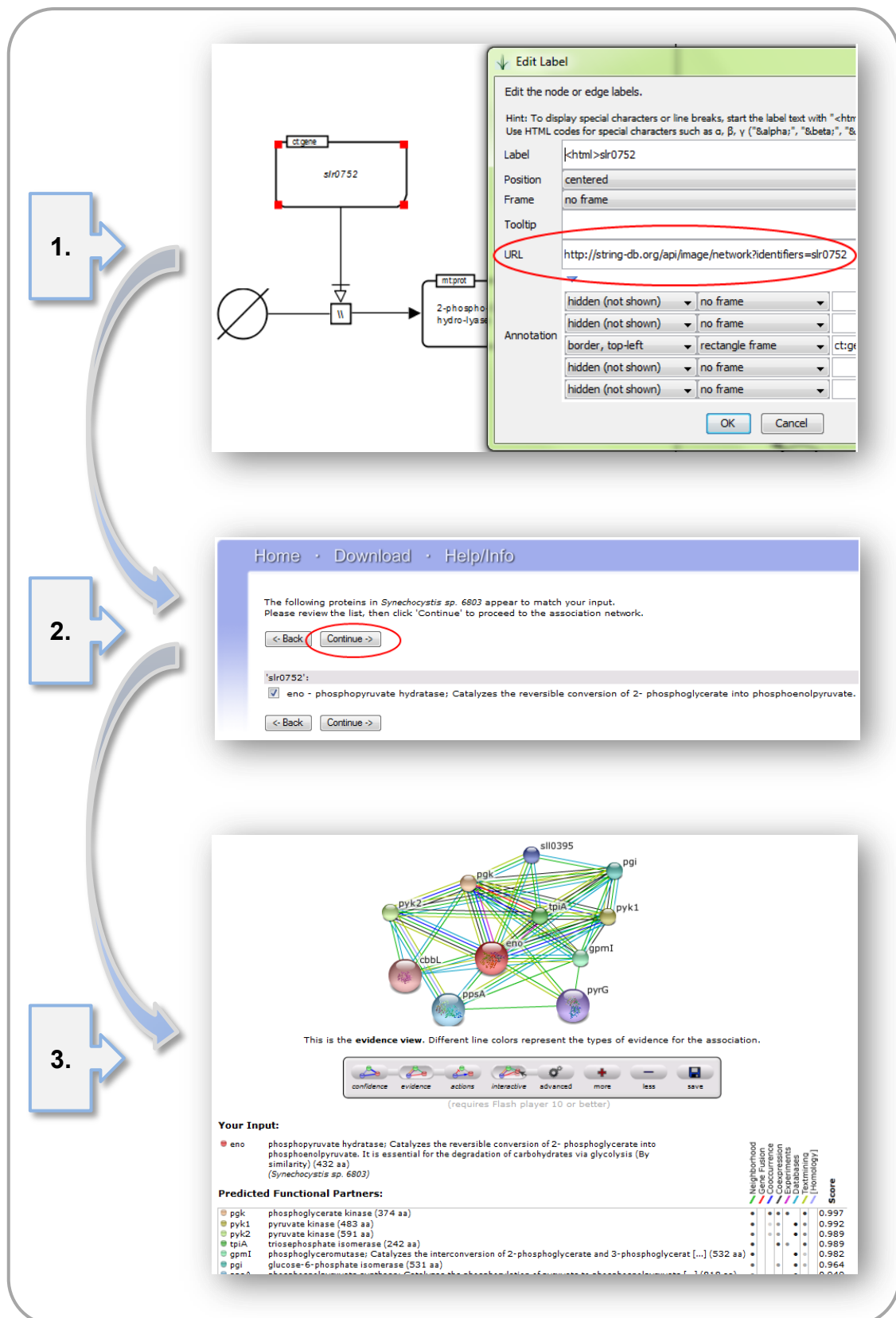
**5. [parameter]  $\triangleq$  identifiers**

Angabe der Parameter, es soll nach definierten Kennungen gesucht werden.

**6. [value]  $\triangleq$  slr0752**

Angabe des Wertes, hier ORF *slr0752*.

Abbildung 20 beschreibt ein Flussdiagramm vom Beginn des Einfügens eines URI in einem metabolischen Modell bis hin zur Ausgabe des Protein-Protein-Interaktionsnetzwerks im Browser. Die 1. Illustration beschreibt das Einfügen des manuell erstellten URI für jeden einzelnen ORF in VANTED. Der Link wird im vorgesehenen Kästchen „URL“ eingefügt und ist zur Hervorhebung rot eingerahmt. Er macht es nun möglich, dass bei einem Klick auf das Bildzeichen des Gens sofort die hinterlegte Adresse im Browser gesucht wird. Diese Eigenschaft ist selbstverständlich nur möglich, wenn die metabolische Karte im Internet veröffentlicht ist. Zunächst gelangt man nun auf eine Seite, welche die Eingabe bestätigt und den gesuchten ORF noch einmal anzeigt (2. Abbildung). Diese Seite bestätigt man mit einem der beiden „Continue“-Button, wobei nachfolgend nur der obere Button markiert ist. Der abschließende Schritt visualisiert das in der STRING Datenbank abgelegte dynamische Netzwerk der Protein-Protein-Interaktionen des gewünschten ORF. Auf dieser Seite kann man außerdem prüfen, woher die Daten der Interaktionen stammen und anschließend gewünschte Metainformationen für nachfolgende Analysen protokollieren.

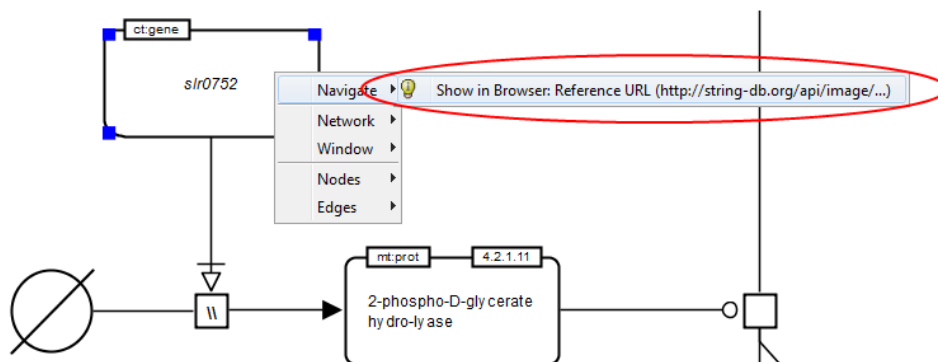


**Abbildung 20: Flussdiagramm - interaktive Verlinkung eines ORF im metabolischen Modell zu dem zugehörigen Protein-Protein-Interaktionsnetzwerk in der STRING-Datenbank**

1. Einfügen des manuell erstellten URI für ein ORF in VANTED 2. Zwischenschritt der Bestätigung 3. Visualisierung des Netzwerks zu potentiellen Interaktionspartnern im Browser

Die Angabe des **[parameter]** wurde bewusst mit **identifiers** gewählt, auch wenn die alleinige Abfrage nach den Interaktionen eines ORF abgerufen wird. Wenn man nur **identifier** als **[parameter]** wählt, bekommt man als Rückgabewert nur das statische Bild der Interaktionen zum Anfrage-ORF und kann keine Metainformationen abrufen, was jedoch nicht sinnvoll ist. Bei dieser Vorgehensweise ist der Zwischenschritt aus der 2. Abbildung nicht nötig, da keine Abfrage der korrekten Auswahl von ORFs erfolgt.

Wenn man sich ein Modell direkt in VANTED anschaut und man trotzdem die Metainformationen aus den Verlinkungen beziehen möchte, muss man einen Rechtsklick auf das entsprechende Bildzeichen ausführen und anschließend über den Punkt „Navigate“ im Drop-down-Menü die „Reference URL“ auswählen und mit einem linken Mausklick bestätigen. Anschließend führt man die Analysen gemäß der 2. Abbildung im Flussdiagramm weiter. Abbildung 21 visualisiert dieses Vorgehen.



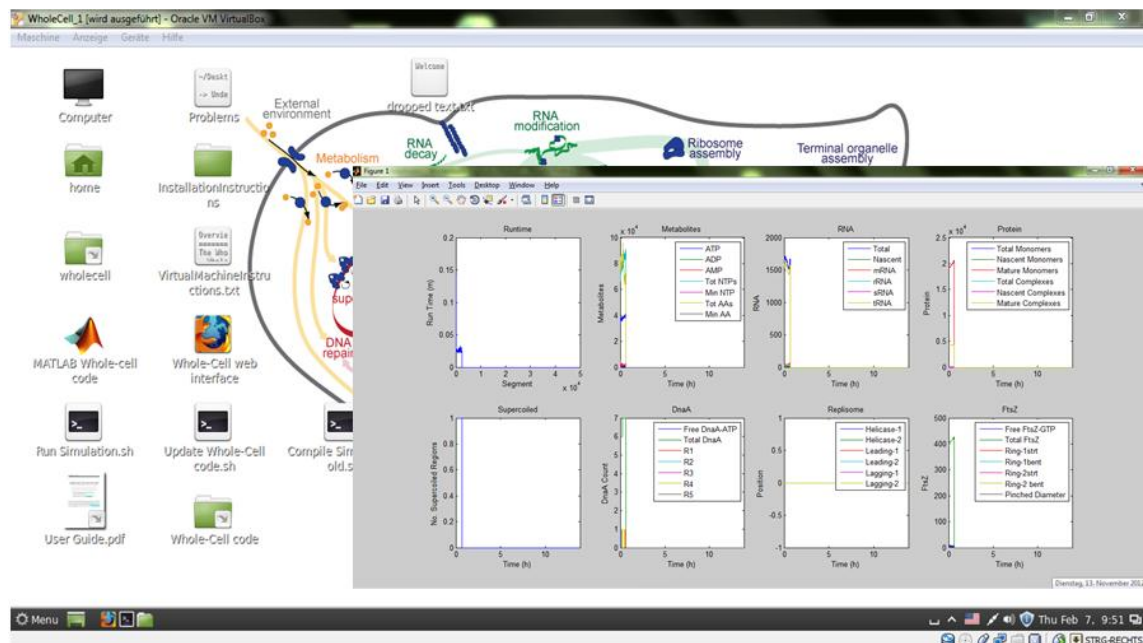
**Abbildung 21: Zugriff auf hinterlegte Links in VANTED**

Um direkt in VANTED Zugriff auf hinterlegte Links ins Internet zu öffnen genügen ein Rechtsklick auf das jeweilige Bildzeichen sowie die Bestätigung der „Reference URL“ unter dem Punkt „Navigate“ im Drop-down-Menü.

### 3.4 KNOWLEDGEBASE FÜR CYANOBAKTERIEN

Das System zur Datenverwaltung und Simulation metabolischer Umsätze von *Mycoplasma genitalium* wurde erfolgreich als lokales System installiert. Der erste Versuch der Installation des Quellcodes auf einer mittels VirtualBox eingerichteten virtuellen Maschine mit dem Betriebssystem Ubuntu, Version 12.04.1 LTS, scheiterte [URL-44]. Die Knowledgebase ist schlicht zu umfangreich und individuell, um sie ohne größere Probleme auf jedem System zu installieren und

einzurichten. Nach Rücksprache mit dem Ansprechpartner wurde eine komfortablere Lösung für dieses Problem gesucht, denn vorhandene Komplikationen waren bekannt. Es manifestierte sich das Vorhaben, das gesamte System auf einer individuell angepassten virtuellen Maschine zu implementieren und dadurch eine einheitliche Arbeitsplattform zu schaffen. Folglich wurde diese Idee vom leitenden Entwickler der *Knowledgebase* umgesetzt. Die Daten und Simulationsprozeduren wurden auf einer virtuellen Maschine mit einem Linux Mint 14 Betriebssystem initialisiert und in VirtualBox importiert. Weitere Anpassungen wurden manuell vorgenommen. In Abbildung 22 wird die Arbeitsoberfläche dargestellt sowie der Output einer Simulation metabolischer Umsätze von *Mycoplasma genitalium*.



**Abbildung 22: Arbeitsoberfläche Knowledgebase**

Die Darstellung visualisiert die Arbeitsoberfläche der Knowledgebase mittels des Mint-Betriebssystems und das Ergebnis einer Simulation metabolischer Umsetzungen von dem Organismus *Mycoplasma genitalium*.

Ein artspezifisches zentrales Datenverwaltungssystem wie dieses liefert enorme Vorteile für das globale Verständnis eines Organismus. Umfassende Simulationen können bei Vorlage bzw. Kenntnis praktischer Messwerte unter adaptierten Bedingungen durchgeführt werden und so die Möglichkeit der Interpretation von inter- und intrazellulären Vorgängen fördern. Natürlich können

auch nur ausgewählte Teile des Metabolismus simuliert werden. Allerdings sollte man sich vor der Etablierung eines solchen Vorhabens den Nutzen und Aufwand abwägen. Der zeitliche Aufwand ist als Einzelperson nicht zu stemmen, denn um jegliche Publikationen für eine Spezies durchzuschauen und wichtige Inhalte zu extrahieren ist schlicht ein gesamtes Team erforderlich. Für die Knowledgebase zu *Mycoplasma genitalium* wurden über 900 Publikationen analysiert. Diese Anzahl korreliert etwa mit jener von 835 gelisteten in GoPubMed. Das ist eine semantische Suchmaschine basierend auf PubMed, welche angeforderte Daten direkt in dedizierte Kategorien klassifiziert und somit eine übersichtlichere Literaturrecherche ermöglicht als die normale Suche über PubMed [URL-14]. Dadurch könnte man innerhalb einer Arbeitsgruppe auch direkt verschiedene Aufgabenbereiche verteilen und somit ein transparentes Arbeitsumfeld ermöglichen. Wenn man die Anzahl an Publikationen für *Synechocystis* sp. PCC 6803 vergleichsweise betrachtet, erkennt man deutlich den Mehraufwand, welchen bereits dieser Organismus erfordert. Die Suche liefert im Vergleich zu *Mycoplasma* etwa die doppelte Anzahl mit 1700 Treffern [URL-13]. Legt man eine Bearbeitungszeit von einer Stunde pro Publikation aus, so würde für eine einzelne Person bei einer Tagesarbeitszeit von 8 h eine Dauer von etwa 212 Tagen resultieren. Darüber hinaus wird man auf längere Zeit diesen Schnitt nicht halten können und zwischendurch andere Arbeiten durchführen. Wenn man bedenkt, dass auch noch allgemeine Publikationen gelesen werden müssen, so wird eindeutig, dass bereits die Literaturrecherche den Arbeitsaufwand für eine Person sprengt und eine Arbeitsgruppe etabliert werden müsste. Dies hat außerdem die sinnvolle Konsequenz, dass jedes Mitglied Spezialwissen in bestimmten Bereichen ausprägt und somit effektiver am konsekutiven ComputermodeLL mitarbeiten kann. Ein weiterer Aspekt, welcher Beachtung finden sollte ist die Abschätzung der verarbeiteten zellulären Prozesse. Während das Genom von *Mycoplasma genitalium* nur eine Größe von etwa 580 kb umfasst, ist jenes von *Synechocystis* sp. PCC 6803 ca. 350000 kb groß [Fraser *et al.*, 1995; URL-8]. Noch deutlicher wird dieser theoretische Ansatz beim Vergleich der Anzahl an identifizierten Genen. In *Mycoplasma* wurden 525 Gene detektiert, in *Synechocystis* dagegen 3317 [Karr *et al.*, 2012; URL-8]. Die Diversität zellulärer Prozesse wird in *Synechocystis* deswegen

weitaus größer und komplexer sein. Auch technische Hindernisse gilt es zu überbrücken, denn für solche komplexen Simulationen sollten große Rechencluster verwendet werden, damit sich die Berechnungszeiten in einem tolerablen Bereich befinden.

Wenn man eine vergleichbare Arbeit initiieren würde, so müsste man konsequent Vor- und Nachteile diskutieren. Außerdem benötigt man unter idealen Umständen ein Team von Wissenschaftlern aus den Bereichen Biologie, Bioinformatik, Informatik, Mathematik sowie Physik, um die Aufgaben sinnvoll aufzuteilen und alle Daten logisch und strukturiert aufzuarbeiten. Ein solches Projekt eröffnet aber auch viele Möglichkeiten der Erforschung bisher unbekannter zellulärer Abläufe oder jene, welche sich unter Laborbedingungen nur schwer realisieren lassen.

### **3.5 VERGLEICH VON VISUALISIERUNGSSOFTWARE FÜR DIE DARSTELLUNG BIOLOGISCHER UND METABOLISCHER ABLÄUFE**

Die Analyse der 14 getesteten bioinformatischen Softwaretools für die Visualisierung metabolischer Netzwerke wurde bezugnehmend der

- Installationsprozedur;
- Anwendungsmöglichkeiten/Komplexität;
- Benutzerfreundlichkeit während des Visualisierungsprozesses

durchgeführt.

Die Einschätzung dieser Punkte unterliegt sicherlich teilweise subjektiven Aspekten. Eine objektive Einschätzung kann bereits durch die Diversität der Anwendungsmöglichkeiten nicht ohne weiteres gewährleistet werden. Außerdem sollte man wahrscheinlich alle oder nahezu alle auf dem Markt befindliche Software auf diesem Gebiet kennen, um sich anzumaßen, eine zuverlässige objektive Einschätzung geben zu können.

## INSTALLATIONSPROZEDUR

Während des Installationsvorgangs zeigten die Programme bereits Unterschiede betreffend den Aufwand und der Benutzerfreundlichkeit. Die meisten Programme waren jedoch durch ein übersichtliches und intuitives Setup sehr schnell und unkompliziert zu installieren. Probleme oder eine Falschinstallation sind nahezu ausgeschlossen. Negativbeispiel war jedoch die Software Pathway Logic. Diese konnte trotz intensiver Versuche nicht erfolgreich installiert und demzufolge auch nicht getestet werden. Im Anhang befinden sich für alle Tools außer Pathway Logic, bei dem die Installation wie bereits geschildert scheiterte, Abbildungen deren Arbeitsoberflächen. Einige Programme sind auch für die akademische Nutzung nur für einen zeitlich begrenzten Zeitraum verfügbar, für weiterführende Arbeiten ist eine kostenpflichtige Lizenz von Nöten.

## ANWENDUNGSMÖGLICHKEITEN/KOMPLEXITÄT

Die Analysen zeigten, dass nicht alle Programme dieselbe Fülle an Möglichkeiten der Editierung und Erstellung von Netzwerken aufweisen. Das ist verständlich, denn die Biologie hält eine Vielzahl unterschiedlicher Mechanismen bereit. Das Tool Pathway Projector ist das einzige im Studienrahmen, welches webbasiert im Browser ausgeführt wurde. Auch eine lokale Version ist verfügbar, jedoch bietet diese keine Vorteile. Es ist die einzige Software, welche nicht den Export der Arbeiten als Bild unterstützt, stattdessen kann man diese nur als XML-File lokal speichern. Viele Programme unterstützen den Import von experimentell ermittelten Daten, sodass ein Mapping auf ein importiertes oder erstelltes Netzwerk realisiert werden kann. Tools mit dieser außerordentlich wichtigen und praktischen Eigenschaft sind bspw. VANTED, Cytoscape, Omix und Cell Designer. Interessant ist in diesem Zusammenhang auch die Netzwerkanalyse, bspw. zur Bestimmung des kürzesten Pfades zwischen zwei Metaboliten oder die Anwendung statistischer Funktionen. Analysen dieser Art sind u.a. mit VANTED, Cytoscape sowie Ondex realisierbar. Die Analyse von Netzwerken könnte selbstverständlich durch eigens implementierte Programme geschehen. Dazu muss die Software allerdings über eine Programmierschnittstelle (*engl.* API) verfügen. Diese Eigenschaft weisen mitt-



lerweile aber fast alle getesteten Tools auf. Auch die Anzeige bzw. der Verweis zu Metadaten kann Analysen und Visualisierungen vereinfachen, unterstützt u.a. von VANTED, Cell Designer und Cytoscape. Abschließend soll erwähnt werden, dass auch die Unterstützung von SBGN immer häufiger Betrachtung findet. Tools, welche diese Notation u.a. unterstützen sind VANTED, BioUML, Cytoscape und PathwayLab. Nachfolgend beschreibt Tabelle 10 die relevanten Eigenschaften zusammenfassend für die getesteten Programme:

**Tabelle 10: Übersicht der Möglichkeiten getesteter Visualisierungsprogramme**

Die Tabelle charakterisiert die grundsätzlichen Eigenschaften der während der Studie getesteten Visualisierungssoftware. Weist ein Programm die geprüfte Eigenschaft auf, so ist diese mit einem Häkchen gekennzeichnet.

	BINA	BIO TAPESTRY	BIOUML	CELL DESIGNER	CYTOSCAPE	MAPMAN	OMIX	ONDEX	PATHVISIO	PATHWAY LOGIC	PATHWAYLAB	PATHWAY PROJECTOR	VANTED	VISANT
<b>Bildexport</b>	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
<b>Import experimenteller Daten</b>	✓	✓			✓	✓	✓		✓		✓	✓	✓	✓
<b>Mapping von Daten</b>	✓	✓		✓	✓	✓	✓	✓	✓			✓	✓	
<b>Netzwerkanalyse</b>	✓				✓			✓		✓			✓	✓
<b>Verfügbarkeit statistischer Funktionen</b>					✓			✓	✓		✓		✓	
<b>Programmierschnittstelle</b>		✓	✓	✓	✓	✓	✓	✓	✓		✓		✓	✓
<b>Zugriff auf Metadaten</b>	✓		✓		✓			✓	✓			✓	✓	✓
<b>Unterstützung von SBGN</b>			✓	✓	✓				✓				✓	

Die Programme sollten je nach Art der gewünschten Anwendung individuell ausgewählt und eingesetzt werden. Wenn bspw. ausschließlich genregulatorische Netzwerke angelegt werden sollen kann sehr gut das Tool BioTapestry genutzt werden. Die beiden Programme VANTED und Cytoscape haben sich bei der Analyse jedoch als umfangreichste etabliert.

### **BENUTZERFREUNDLICHKEIT WÄHREND DES VISUALISIERUNGSPROZESSES**

Hinsichtlich der Benutzerfreundlichkeit gibt es große Unterschiede, denn was nützt es ein Tool mit vielen Funktionen und ausgereiften Details zu besitzen, wenn diese zu versteckt für den Anwender im Programmlayout implementiert sind. Jedoch ist besonders die Einschätzung der Benutzerfreundlichkeit eine in vielen Aspekten subjektive Entscheidung. Dabei kommt es auf die Erfahrung des Users mit Software, insbesondere bioinformatischer, an. Jedes Tool bedarf einer gewissen Einarbeitungszeit. Während der Testphase haben sich besonders VANTED und Omix als unkompliziert im Umgang erwiesen. Aus diesem Grund wurde VANTED auch als Visualisierungssoftware ausgewählt.

## **3.6 AUSWERTUNG DER STUDIE ÜBER DIE AKZEPTANZ EINES VISUALISIERUNGSSTANDARDS IN DER BIOLOGIE**

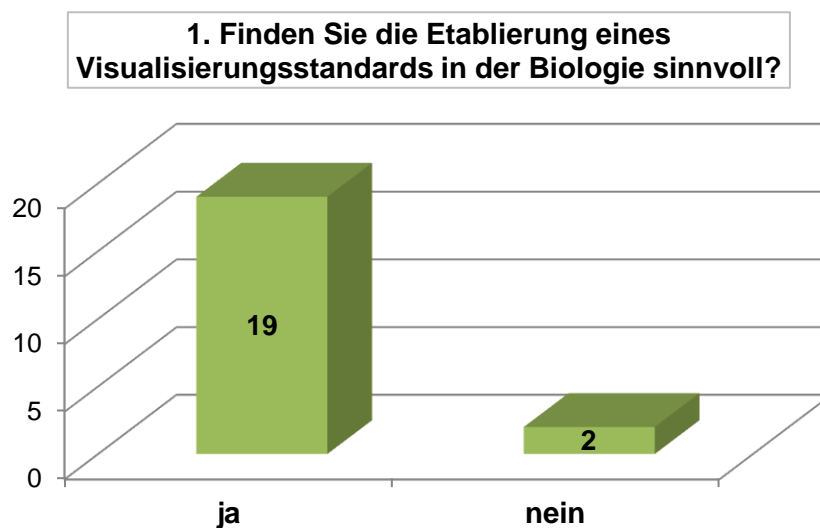
Die Studie über die Evaluierung der Akzeptanz eines Visualisierungsstandards in der Biologie wurde über einen Befragungszeitraum von 77 Tagen realisiert. Insgesamt wurden über 300 Anfragen an Biowissenschaftler in ganz Deutschland sowie an einige ausgewählte ausländische Probanden versendet. Die prozentuale Teilnahme an dieser Studie war trotz des Engagements ernüchternd und lag nur bei ca. 7 %. Umfassend wurden 18 deutschsprachige sowie 3 englischsprachige Fragebögen ausgefüllt. Eine Aufgliederung des genauen Umfrageergebnisses verdeutlicht Tabelle 11.

**Tabelle 11: Quantitative Auflistung des Teilnahmeergebnisses der Studie über die Evaluierung der Akzeptanz eines Visualisierungsstandards in der Biologie**

Die Studie wurde sowohl in deutscher als auch in englischer Sprache verfasst, um auch ausländische Probanden berücksichtigen zu können. Insgesamt wurden beide Fragebögen 63-mal während des Umfragezeitraums angeklickt und 21 Fragebögen, 18 in Deutsch und 3 in Englisch, ausgefüllt.

Fragebogen	ausgefüllte Fragebögen	Klicks
deutschsprachig	18	47
englischsprachig	3	18
<b>gesamt</b>	<b>21</b>	<b>65</b>

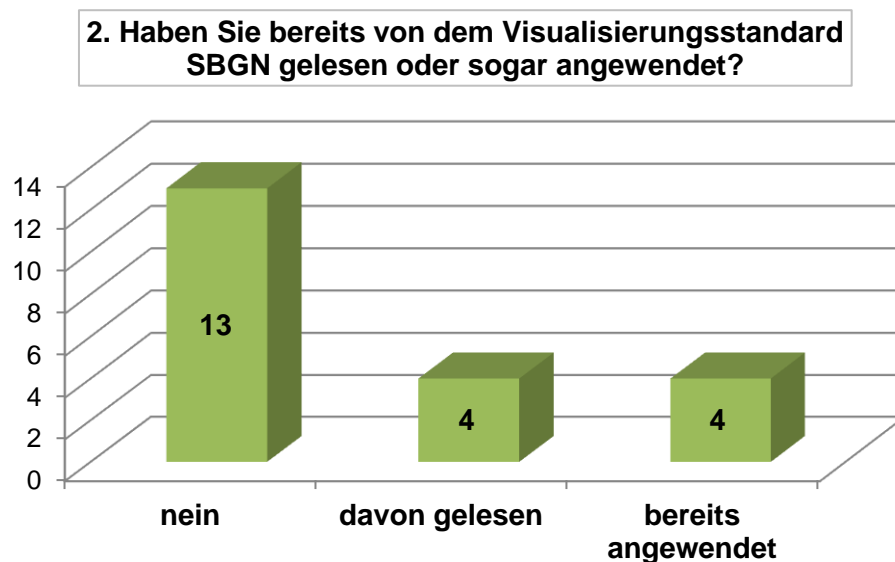
Alle Fragebögen wurden vollständig ausgefüllt. Die Umfrage lässt den Schluss zu, dass die meisten Teilnehmer die Intention der Etablierung eines Visualisierungsstandards in der Biologie sinnvoll finden, denn 19 Probanden haben auf die erste Frage bejahend geantwortet, dargestellt in Abbildung 23.



**Abbildung 23: Verteilung der Antworten auf die Frage „Finden Sie die Etablierung eines Visualisierungsstandards in der Biologie sinnvoll?“**

Die Etablierung eines Visualisierungsstandards in der Biologie fanden 19 Probanden sinnvoll, 2 Teilnehmer sahen das nicht so.

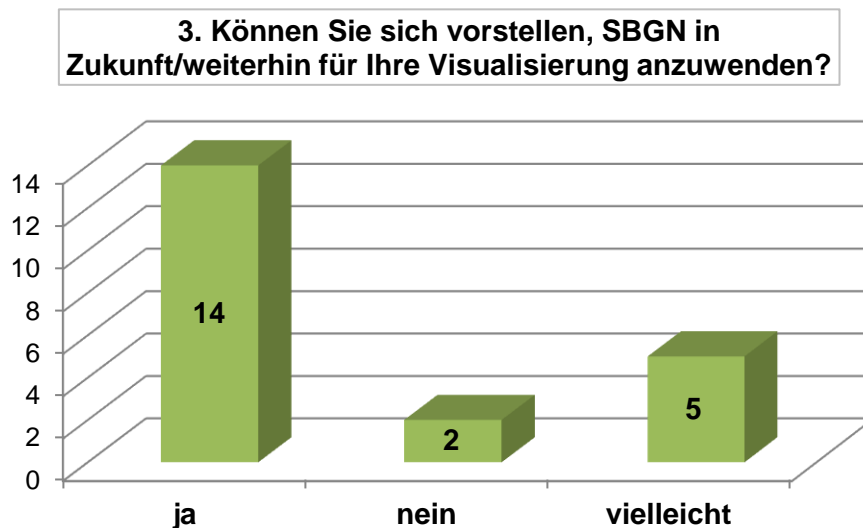
Von der *Systems Biology Graphical Notation* haben jedoch mehr als 60 % der Probanden noch nichts gelesen, noch weniger haben diese selbst schon bei Visualisierungen angewendet. Dargestellt ist diese Verteilung in Abbildung 24.



**Abbildung 24:** Verteilung der Antworten auf die Frage „Haben Sie bereits von dem Visualisierungsstandard SBGN gelesen oder sogar angewendet?“

*Den Visualisierungsstandard SBGN haben bisher 4 Teilnehmer der Studie angewendet und 4 weitere haben von dieser Notation gelesen. Die verbleibenden 13 Probanden haben noch nichts von dieser Thematik gehört.*

Im Gegensatz dazu könnten sich aber fast alle Teilnehmer vorstellen, diese Notation in Zukunft anzuwenden, wie es in Abbildung 25 ersichtlich wird.



**Abbildung 25: Verteilung der Antworten auf die Frage „Können Sie sich vorstellen, SBGN in Zukunft/weiterhin für Ihre Visualisierung anzuwenden?“**

*Ein Anteil von 66 % der Probanden kann sich definitiv vorstellen, SBGN zu nutzen. Lediglich 2 Teilnehmer schließen eine Anwendung konsequent aus, 5 Probanden stellen eine Nutzung in Aussicht.*

Die Auswertung von Frage 4 der Studie, beziehend der Ansprüche an einen Visualisierungsstandard, hat gezeigt, dass sowohl lobende als auch kritische Worte von den Teilnehmern gefunden wurden. Das Erstellen von Netzwerken wird als einfach und umfassend beschrieben, doch sehen einige diese Darstellungsform als „zu technisch anmutende Notation“ an und prognostizieren, dass Biologen dies nicht anerkennen werden. Eine Auswahl dieser Antworten ist in Tabelle 12 dargestellt, alle übrigen Antworten sind im Anhang in Tabelle 15 vermerkt. Dabei ist zu berücksichtigen, dass einige Kommentare teilweise auf Unkenntnis basieren müssen, denn sowohl die Anwendung von Algorithmen für die Layout-Optimierung als auch die kreisförmige Anordnung von Stoffwechselwegen ist möglich, insbesondere in Verbindung mit der Visualisierungssoftware VANTED.

**Tabelle 12: Darstellung einer Auswahl von Antworten auf die Frage „Stellen Sie explizite Ansprüche an einen Visualisierungsstandard? Gern können Sie nachfolgend einige Ihrer Wünsche äußern.“**

*Das Abgeben von Antworten wurde von 9 Personen praktiziert. Dabei wurden sowohl lobende als auch kritische Worte geäußert.*

4. Stellen Sie explizite Ansprüche an einen Visualisierungsstandard? Gern können Sie nachfolgend einige Ihrer Wünsche äußern.
einfach und verständlich
umfassend in der Repräsentation des Wissens
Standards sollten die Prozesse aus unterschiedlichen Zyklen untereinander vergleichbar machen.
er muss von den experimentell arbeitenden Biologen auf breiter Front akzeptiert werden
Stoffwechselwege müssen so visualisiert sein, dass Biologen mit den Diagrammen etwas anfangen können
SBGN ist eine zu technisch anmutende Notation, die einer dynamischen, generischen Darstellungsweise, wie es von Biologen bevorzugt wird, entgegensteht

Die Etablierung eines Visualisierungsstandards in der Biologie liefert eine Diskussionsgrundlage, welche von verschiedenen Seiten betrachtet werden sollte. Zum einen besteht die reelle Chance alle oder zumindest einen Großteil biologischer Vorgänge mit einheitlichen Bildzeichen graphisch darzustellen. Kritiker könnten darauf beharren und eine solche Notation als unzureichend klassifizieren, falls es sich abzeichnet, dass eben nicht alle Prozesse einheitlich abzubilden sind. Befürworter mögen sagen, dass es ein Schritt in eine vernünftige und richtige Richtung ist. Die Art der graphischen Abbildung von Netzwerken mittels der *Systems Biology Graphical Notation* ist für einen Bioinformatiker klar verständlich und intuitiv, da viele Softwaretools eine vergleichbare Darstellungsform bieten. Dieser Visualisierungsstandard sollte weiter in Arbeiten publiziert werden, um diese Darstellungsform einer größeren Leserschaft zu präsentieren und klar die Vorzüge aufzuzeigen.

## 4 ZUSAMMENFASSUNG

Mittels bioinformatischer Visualisierungsmethoden wurde in der vorliegenden Arbeit der Wasserstoffmetabolismus der Cyanobakterien *Synechocystis* sp. PCC 6803 und *Nostoc* sp. PCC 7120 aufgearbeitet. Dazu wurden genomische Informationen sowie theoretische und experimentelle Charakteristika zu den relevanten Proteinkomplexen Nitrogenase sowie bidirektionale und uptake Hydrogenase mittels einer intensiven Literaturrecherche ermittelt. Außerdem wurde existierende bioinformatische Visualisierungssoftware im Rahmen einer Studie getestet und VANTED in Verbindung mit der *Systems Biology Graphical Notation* als Darstellungsplattform gewählt, um infolgedessen das evaluierte metabolische Modell des Wasserstoffmetabolismus von *Synechocystis* sp. PCC 6803 darzustellen. Durch diese Art der Präsentation wird ein hohes Maß an Modularität gewährleistet und die effektive Einbindung von Metainformationen in Form von putativen Promoter-Bindestellen und interaktiven Verlinkungen ins Internet konnte realisiert werden. Im Modell visualisierte Transkriptionsfaktoren wurden in der grundlegenden Literaturrecherche sowie durch die Analyse existierender Operon-Datenbanken identifiziert. Nach erfolgreicher Umsetzung wurde dieses praktische Vorgehen auch für den Organismus *Nostoc* sp. PCC 7120 durchgeführt und ein adaptiertes metabolisches Modell seines Wasserstoffmetabolismus entwickelt. Die Modelle sind auf der Webseite <http://www.hs-mittweida.de/wuenschi/SBGN/> zugänglich.

Aufgrund der Tatsache, dass noch nie eine Studie über die Toleranz eines Visualisierungsstandards in der Biologie durchgeführt wurde, war es Ansporn genug, dieses in der vorliegenden Arbeit anzustellen. Dazu wurden über 300 Biowissenschaftler zu der Teilnahme an dieser Studie motiviert. Resümierend lässt sich formulieren, dass der Großteil der Probanden sich nicht vor einem Visualisierungsstandard verschließt, jedoch wird angezweifelt, ob Biologen diese Darstellungsform langfristig akzeptieren werden.

Abschließend wurde die Etablierung einer *Knowledgebase* für Cyanobakterien erörtert. Diese Daten- und Simulationsplattform gewährleistet die Verwaltung von relevanten Informationen und bietet neben dem großen Aufwand der Zentralisierung aller Daten die Möglichkeit, bisher wenig verstandene intra- und interzelluläre Prozesse zu detektieren und aufzuklären.

## 5 AUSBLICK

Das Themengebiet der Datenvisualisierung in der Biologie mittels standardisierter Prozeduren ist noch nicht sehr ausgebaut, bietet aber die Chance einer umfassenden Darstellungsform biologischer und biochemischer Abläufe. Dagegen könnten künftig weitere intensive Bemühungen und Arbeiten stattfinden. Ein, wenngleich sehr aufwendiger, Schritt zur Erreichung eines hohen Maßes an Universalität der Visualisierung metabolischer Abläufe könnte die umfassende Darstellung der Boehringer-Karte sein. Dadurch könnte man zukünftige Arbeiten von Wissenschaftlern erleichtern, würde SBGN in einem hohen Maß publizieren und entscheidend zu der Intention der Verbreitung eines weltweiten Visualisierungsstandards beitragen. Dies würde aber auch prozedurale Veränderungen bei der Darstellung verlangen. Ein eigens für SBGN implementierter Editor mit weiteren Möglichkeiten, wie der dreidimensionalen Darstellung von Netzwerken, der aktiven Einbeziehung von Metainformation wie bspw. Proteinstrukturen sowie die Verknüpfung von Netzwerkreferenzen zur Darstellung unterschiedlicher Ebenen im Visualisierungsprozess würde einen hohen Standard signalisieren und SBGN in seiner Etablierung wahrscheinlich voran bringen. Wie die im Umfang dieser Arbeit durchgeführte Studie gezeigt hat, sehen viele noch nicht den Mehrwert in dieser standardisierten Darstellungsform. Wenn man eine optisch versiertere und intuitive Visualisierung schafft, kann das Konzept gelingen.

Auch weiterführende Arbeiten und die Etablierung der Darstellung und Simulation metabolischer Abläufe in Form organismenspezifischer Data-Warehouse-Systemen wie der *Knowledgebase* für *Mycoplasma genitalium* sollten für Modellorganismen wie *Synechocystis* sp. PCC 6803 entwickelt werden. Sie bieten die Möglichkeit des zentralen Abrufs von Informationen und beinhalten gesammeltes Wissen von Jahrzehnten Forschung.

Ein Trend auf dem Gebiet der Visualisierung biologischer Netzwerke mittels bioinformatischer Methoden ist außerdem die Etablierung des modularen Aufbaus, d.h. vorgefertigte Module eines Modells oder vollständige Modelle werden bereitgestellt und der Anwender muss die Module nur nach seinen Anforderungen beschriften und kann verschiedene zu einem gesamten Modell zusammenfassen. SBGN-Bricks beschreibt diesen Weg [Junker *et al.*, 2012]. Die



BioModels Database des EBI verwaltet hunderte Modelle biochemischer Prozesse, welche zum Download in verschiedenen Formaten zur Verfügung stehen und so als Grundlage komplexer Netzwerke und als Grundlage eigener Forschungen verwendet werden können [URL-45]. Weiterhin werden Arbeitsoberflächen wie die *Biological Integrated Knowledge Environment* (BioBIKE) etabliert, um auch Biowissenschaftlern mit geringeren informatischen Kenntnissen komplexere bioinformatische Analysen zu ermöglichen [Elhai *et al.*, 2009; URL-46].

## LITERATURVERZEICHNIS

- Agervald, Stensjö, Holmqvist & Lindblad (2008) Transcription of the extended *hyp*-operon in *Nostoc* sp. strain PCC 7120. *BMC Microbiology* 8:69
- Agervald, Zhang, Stensjo, Devine & Lindblad (2010) CalA, a Cyanobacterial AbrB Protein, Interacts with the Upstream Region of *hypC* and Acts as a Repressor of Its Transcription in the Cyanobacterium *Nostoc* sp. Strain PCC 7120. *Applied and Environmental Microbiology* 76:880-890
- Anderson & McIntosh (1991) Light-Activated Heterotrophic Growth of the Cyanobacterium *Synechocystis* sp. Strain PCC 6803: a Blue-Light-Requiring Process. *Journal of Bacteriology* 173:2761-2767
- Baebprasert, Lindblad & Incharoensakdi (2010) Response of H<sub>2</sub> production and Hox-hydrogenase activity to external factors in the unicellular cyanobacterium *Synechocystis* sp. strain PCC 6803. *International Journal of Hydrogen Energy* 35:6611-6616
- Bandyopadhyay, Stöckel, Min, Sherman & Pakrasi (2010) High rates of photo-biological H<sub>2</sub> production by a cyanobacterium under aerobic conditions. *Nature Communications* 1:1-7
- Beltrame, Calura, Popovici, Rizzetto, Rivero Guedez, Donato, Romualdi, Draghici & Cavalieri (2011) The Biological Connection Markup Language: a SBGN-compliant format for visualization, filtering and analysis of biological pathways. *Bioinformatics* 27:2127-2133
- Bothe, Schmitz, Yates & Newton (2010) Nitrogen Fixation and Hydrogen Metabolism in Cyanobacteria. *Microbiology and Molecular Biology Reviews* 74:529-551
- Bruggeman & Westerhoff (2006) The nature of systems biology. *Trends in Microbiology* 15:45-50
- Bry & Kröger (2003) A computational biology database digest: data, data analysis, and data management. *Distributed and Parallel Databases* 13:7-42
- Burrows, Chaplen & Ely (2008) Optimization of media nutrient composition for increased photofermentative hydrogen production by *Synechocystis* sp. PCC 6803. *International Journal of Hydrogen Energy* 33:6092-6099

- Carrasco, Buetitner & Golden (1995) Programed DNA rearrangement of a cyanobacterial *hupL* gene in heterocysts. Proceedings of the National Academy of Sciences of the United States of America 92:791-795
- Carrasco, Holliday, Hansel, Lindblad & Golden (2005) Heterocyst-Specific Excision of the *Anabaena* sp. Strain PCC 7120 *hupL* Element Requires *xisC*. Journal of Bacteriology 187:6031-6038
- Carrasco, Ramaswamy, Ramasubramanian & Golden (1994) *Anabaena xisF* gene encodes a developmentally regulated site-specific recombinase. Genes Dev. 8:74-83
- Caspi, Altman, Dreher, Fulcher, Subhraveti, Keseler, Kothari, Krummenacker, Latendresse, Mueller, Ong, Paley, Pujar, Shearer, Travers, Weerasinghe, Zhang & Karp (2011) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Research 40:742-753
- Czauderna, Klukas & Schreiber (2010) Editing, validating and translating of SBGN maps. Bioinformatics 26:2340-2341
- Dasa & Veziroglu (2001) Hydrogen production by biological processes: a survey of literature. International Journal of Hydrogen Energy 26:13-28
- Dutheil, Saenkham, Sakr, Leplat, Ortega-Ramos, Bottin, Cournac, Cassier-Chauvat & Chauvat (2012) The AbrB2 Autorepressor, Expressed from an Atypical Promoter, Represses the Hydrogenase Operon To Regulate Hydrogen Production in *Synechocystis* Strain PCC6803. Journal of Bacteriology 194:5423-5433
- Dutta, De, Chaudhuri & Bhattacharya (2005) Hydrogen production by Cyanobacteria. BMC Microbial Cell Factories 4:36
- Edwards, Kuznetsov & David (2007) Hydrogen energy. Philosophical Transactions A 365:1043-1056
- Ehira & Ohmori (2011) NrrA, a Nitrogen-regulated Response Regulator Protein, Controls Glycogen Catabolism in the Nitrogen-fixing Cyanobacterium *Anabaena* sp. Strain PCC 7120. The Journal of Biological Chemistry 286:38109-38114
- Elhai, Taton, Massar, Myers, Travers, Casey, Slupesky & Shrager (2009) BioBIKE: A Web-based, programmable, integrated biological knowledge base. Nucleic Acids Research 37:28-32

- Ermolaeva, White & Salzberg (2001) Prediction of operons in microbial genomes. *Nucleic Acids Research* 29:1216-1221
- Finn, Marshall & Bateman (2005) iPfam: visualization of protein-protein interactions in PDB at domain and amino acid resolutions. *Bioinformatics* 21:410-412
- Flores & Herrero (1994) Requirement of the regulatory protein NtcA for the expression of nitrogen assimilation and heterocyst development genes in the cyanobacterium *Anabaena* sp. PCC 7120. *Molecular Microbiology* 14:823-832
- Flores & Herrero (2010) Compartmentalized function through cell differentiation in filamentous cyanobacteria. *Nature Reviews Microbiology* 8:39-50
- Fraser, Gocayne, White, Adams, Clayton, Fleischmann, Bult, Kerlavage, Sutton, Kelley, Fritchman, Weidman, Small, Sandusky, Fuhrmann, Nguyen, Utterback, Saudek, Phillips, Merrick, Tomb, Dougherty, Bott, Hu, Lucier, Peterson, Smith, Hutchison & Venter (1995) The minimal gene complement of *Mycoplasma genitalium*. *Science* 270:397-403
- Fuchs & Schlegel (2007) *Allgemeine Mikrobiologie*. 8. Auflage. Georg Thieme Verlag Stuttgart
- Gutekunst (2006) Regulation der bidirektionalen NiFe-Hydrogenase im Cyanobakterium *Synechocystis* sp. PCC 6803. Christian-Albrechts-Universität zu Kiel. Dissertation
- Gutekunst, Phunpruch, Schwarz, Schuchardt, Schulz-Friedrich & Appel (2005) LexA regulates the bidirectional hydrogenase in the cyanobacterium *Synechocystis* sp. PCC 6803 as a transcription activator. *Molecular Microbiology* 58:810-823
- Gutthann, Egert, Marques & Appel (2006) Inhibition of respiration and nitrate assimilation enhances photohydrogen evolution under low oxygen concentrations in *Synechocystis* sp. PCC 6803. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1767:161-169
- Hoffmann (2008) Charakterisierung der für die Reifung der Hydrogenase und anderer Metall-Enzyme von *Synechocystis* sp. PCC 6803 zuständigen Gene. Christian-Albrechts-Universität zu Kiel. Dissertation
- Hoffmann, Gutekunst, Klissenbauer, Schulz-Friedrich & Appel (2006) Mutagenesis of hydrogenase accessory genes of *Synechocystis* sp. PCC 6803

- Additional homologues of *hypA* and *hypB* are not active in hydrogenase maturation. FEBS Journal 273:4516-4527
- Holmqvist (2010) The Cyanobacterial Uptake Hydrogenase - Regulation, Maturation and Function. Uppsala University. Dissertation
- Jensen, Kuhn, Stark, Chaffron, Creevey, Muller, Doerks, Julien, Roth, Simonovic, Bork & von Mering (2008) STRING 8 - a global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Research 37:412-416
- Junker, Klukas & Schreiber (2006) VANTED: A system for advanced data analysis and visualization in the context of biological networks. BMC Bioinformatics 7:109
- Junker, Sorokin, Czauderna, Schreiber & Mazein (2012) Wiring diagrams in biology: towards the standardized representation of biological information. Trends in Biotechnology 30:555-557
- Kanehisa, Goto, Kawashima, Okuno & Hattori (2004) The KEGG resource for deciphering the genome. Nucleic Acids Research 32:277-280
- Kaneko, Nakamura, Wolk, Kuritz, Sasamoto, Watanabe, Iriguchi, Ishikawa, Kawashima, Kimura, Kishida, Kohara, Matsumoto, Matsuno, Muraki, Nakazaki, Shimpō, Sugimoto, Takazawa, Yamada, Yasuda & Tabata (2001) Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium *Anabaena* sp. Strain PCC 7120. DNA Research 8:205-213
- Kaneko, Sato, Kotani, Tanaka, Asamizu, Nakamura, Miyajima, Hirosawa, Sugiura, Sasamoto, Kimura, Hosouchi, Matsuno, Muraki, Nakazaki, Naruo, Okumura, Shimpō, Takeuchi, Wada, Watanabe, Yamada, Yasuda & Tabata (1996) Sequence Analysis of the Genome of the Unicellular Cyanobacterium *Synechocystis* sp. Strain PCC6803. II. Sequence Determination of the Entire Genome and Assignment of Potential Protein-coding Regions. DNA Research 3:109-136
- Kaneko, Tanaka, Sato, Kotani, Sazuka, Miyajima, Sugiura & Tabata (1995) Sequence Analysis of the Genome of the Unicellular Cyanobacterium *Synechocystis* sp. Strain PCC6803. I. Sequence Features in the 1 MB Region from Map Position 64% to 92% of the Genome. DNA Research 2:153-166
- Kanhere & Bansal (2005) A novel method for prokaryotic promoter prediction based on DNA stability. BMC Bioinformatics 6:1

- Kanhere & Bansal (2005) Structural properties of promoters: similarities and differences between prokaryotes and eukaryotes. *Nucleic Acids Research* 33:3165-3175
- Karp, Ouzounis, Moore-Kochlacs, Goldovsky, Kaipa, Ahrén, Tsoka, Darzentas, Kunin & López-Bigas (2005) Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. *Nucleic Acids Research* 33:6083-6089
- Karr, Sanghvi, Macklin, Gutschow, Jacobs, Bolival, Assad-Garcia, Glass & Covert (2012) A Whole-Cell Computational Model Predicts Phenotype from Genotype. *Cell* 150:389-401
- Kitano (2002) Systems Biology: A Brief Overview. *Science* 295:1662-1664
- Klukas & Schreiber (2010) Integration of -omics data and networks for biomedical research with VANTED. *Journal of Integrative Bioinformatics* 7(2):112
- Klukas, Junker & Schreiber (2006) The VANTED software system for transcriptomics, proteomics and metabolomics analysis. *Journal of Pesticide Science* 31:289-292
- Knoop, Zilliges, Lockau & Steuer (2010) The Metabolic Network of *Synechocystis* sp. PCC 6803: Systemic Properties of Autotrophic Growth. *Plant Physiology* 154:410-422
- Kucho, Okamoto, Tsuchiya, Nomura, Nango, Kanehisa & Ishiura (2005) Global Analysis of Circadian Expression in the Cyanobacterium *Synechocystis* sp. Strain PCC 6803. *Journal of Bacteriology* 187:2190-2199
- Kumar, Mella-Herrera & Golden (2010) Cyanobacterial Heterocysts. *Cold Spring Harbor Perspectives in Biology* 2:1-19
- Le Novère, Hucka, Mi, Moodie, Schreiber, Sorokin, Demir, Wegner, Aladjem, Wimalaratne, Bergman, Gauges, Ghazal, Kawaji, Li, Matsuoka, Villéger, Boyd, Calzone, Courtot, Dogrusoz, Freeman, Funahashi, Ghosh, Jouraku, Kim, Kolpakov, Luna, Sahle, Schmidt, Watterson, Wu, Goryanin, Kell, Sander, Sauro, Snoep, Kohn & Kitano (2009) The Systems Biology Graphical Notation. *Nature Biotechnology* 27:735-741
- Le Novère, Moodie, Sorokin, Schreiber & Mi (2010) Systems Biology Graphical Notation: Entity Relationship language Level 1. *Nature Precedings*

- Lindberg (2003) Cyanobacterial Hydrogen Metabolism - Uptake Hydrogenase and Hydrogen Production by Nitrogenase in Filamentous Cyanobacteria. Uppsala University. Dissertation
- Lopo, Montagud, Navarro, Cunha, Zille, de Cordoba, Moradas-Ferreira, Tamagnini & Urchueguía (2012) Experimental and Modeling Analysis of *Synechocystis* sp. PCC 6803 Growth. Journal of Molecular Microbiology and Biotechnology 22:71-82
- Marcus, Altman-Gueta, Finkler & Gurevitz (2005) Mutagenesis at Two Distinct Phosphate-Binding Sites Unravels Their Differential Roles in Regulation of Rubisco Activation and Catalysis. Journal of Bacteriology 187:4222-4228
- Marques, Barbosa, Jotta, Coelho, Tamagnini & Gouveia (2011) Biohydrogen production by *Anabaena* sp. PCC 7120 wild-type and mutants under different conditions: Light, nickel, propane, carbon dioxide and nitrogen. Biomass and Bioenergy 35:4426-4434
- Masukawa, Inoue, Sakurai, Wolk & Hausinger (2010) Site-Directed Mutagenesis of the *Anabaena* sp. Strain PCC 7120 Nitrogenase Active Site To Increase Photobiological Hydrogen Production. Applied and Environmental Microbiology 76:6741-6750
- Masukawa, Mochimaru & Sakurai (2002) Disruption of the uptake hydrogenase gene, but not of the bidirectional hydrogenase gene, leads to enhanced photobiological hydrogen production by the nitrogen-fixing cyanobacterium *Anabaena* sp. PCC 7120. Applied Microbiology and Biotechnology 58:618-624
- McCarn, Whitaker, Alam, Vrba & Curtis (1988) Genes Encoding the Alpha, Gamma, Delta, and Four F<sub>0</sub> Subunits of ATP Synthase Constitute an Operon in the Cyanobacterium *Anabaena* sp. Strain PCC 7120. Journal of Bacteriology 170:3448-3458
- Meene, Hohmann-Marriott, Vermaas & Roberson (2005) The three-dimensional structure of the cyanobacterium *Synechocystis* sp. PCC 6803. Arch Microbiol 184:259-270
- Mi, Schreiber, Le Novère, Moodie & Sorokin (2009) Systems Biology Graphical Notation: Activity Flow language Level 1. Nature Precedings
- Montagud, Zelezniak, Navarro, Córdoba, Urchueguía & Patil (2011) Flux coupling and transcriptional regulation within the metabolic network of the

- photosynthetic bacterium *Synechocystis* sp. PCC6803. *Biotechnology Journal* 6:330-342
- Moodie, Le Novère, Demir, Mi & Schreiber (2011) Systems Biology Graphical Notation: Process Description language Level 1. *Nature Precedings*
- Moreno-Hagelsieb & Collado-Vides (2002) A powerful non-homology method for the prediction of operons in prokaryotes. *Bioinformatics* 18:329-336
- Navarro, Montagud, Córdoba & Urchueguía (2009) Metabolic flux analysis of the hydrogen production potential in *Synechocystis* sp. PCC6803. *International Journal of Hydrogen Energy* 34:8828-8838
- Ogata, Goto, Sato, Fujibuchi, Bono & Kanehisa (1999) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* 27:29-34
- Okuda & Yoshizawa (2011) ODB: a database for operon organizations, 2011 update. *Nucleic Acids Research* 39:552-555
- Okuda, Katayama, Kawashima, Goto & Kanehisa (2006) ODB: a database of operons accumulating known operons across multiple genomes. *Nucleic Acids Research* 34:358-362
- Oliveira & Lindblad (2005) LexA, a transcription regulator binding in the promoter region of the bidirectional hydrogenase in the cyanobacterium *Synechocystis* sp. PCC 6803. *FEMS Microbiology Letters* 251:59-66
- Oliveira & Lindblad (2007) An AbrB-Like Protein Regulates the Expression of the Bidirectional Hydrogenase in *Synechocystis* sp. Strain PCC 6803. *Journal of Bacteriology* 190:1011-1019
- Olmedo-Verd, Muro-Pastor, Flores & Herrero (2006) Localized Induction of the *ntcA* Regulatory Gene in Developing Heterocysts of *Anabaena* sp. Strain PCC 7120. *Journal of Bacteriology* 188:6694-6699
- Pertea, Ayanbule, Smedinghoff & Salzberg (2008) OperonDB: a comprehensive database of predicted operons in microbial genomes. *Nucleic Acids Research* 37:479-482
- Price, Huang, Alm & Arkin (2005) A novel method for accurate operon predictions in all sequenced prokaryotes. *Nucleic Acids Research* 33:880-892
- Ramaswamy, Carrasco, Fatmat & Golden (1997) Cell-type specificity of the *Anabaena fdxN*-element rearrangement requires *xisH* and *xisI*. *Molecular Microbiology* 23:1241-1249



- Rangannan & Bansal (2007) Identification and annotation of promoter regions in microbial genome sequences on the basis of DNA stability. *Journal of Bio-sciences* 32:851-862
- Rangannan & Bansal (2009) Relative stability of DNA as a generic criterion for promoter prediction: whole genome annotation of microbial genomes with varying nucleotide base composition. *Molecular BioSystems* 5:1758-1769
- Rangannan & Bansal (2010) High-quality annotation of promoter regions for 913 bacterial genomes. *Bioinformatics* 26:3043-3050
- Rangannan & Bansal (2011) PromBase: a web resource for various genomic features and predicted promoters in prokaryotic genomes. *BMC Research Notes* 4:257
- Sachdev, Fairbrother & Deshayes (2003) Exploring Protein-Protein Interactions with Phage Display. *ChemBioChem* 4:14-25
- Sato, Shimoda, Muraki, Kohara, Nakamura & Tabata (2007) A Large-scale Protein-protein Interaction Analysis in *Synechocystis* sp. PCC6803. *DNA Research* 14:207-216
- Schütz, Happe, Troshina, Lindblad, Leitao, Oliveira & Tamagnini (2004) Cyanobacterial H<sub>2</sub> production - a comparative analysis. *Planta* 218:350-359
- Sjöholm, Oliveira & Lindblad (2007) Transcription and Regulation of the Bidirectional Hydrogenase in the Cyanobacterium *Nostoc* sp. Strain PCC 7120. *Applied and Environmental Microbiology* 73:5435-5446
- Song & Park (1997) Organization and Regulation of the D-Xylose Operons in *Escherichia coli* K-12: XylR Acts as a Transcriptional Activator. *Journal of Bacteriology* 179:7025-7032
- Strauch (1995) In Vitro Binding Affinity of the *Bacillus subtilis* AbrB Protein to Six Different DNA Target Regions. *Journal of Bacteriology* 177:4532-4536
- Stricker, Masepohl, Klipp & Böhme (1997) Identification and Characterization of the *nifV-nifZ-nifT* Gene Region from the Filamentous Cyanobacterium *Anabaena* sp. Strain PCC 7120. *Journal of Bacteriology* 179:2930-2937
- Szklarczyk, Franceschini, Kuhn, Simonovic, Roth, Minguéz, Doerks, Stark, Müller, Bork, Jensen & von Mering (2010) The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Research* 39:561-568

- Tamagnini, Axelsson, Lindberg, Oxelfelt, Wünschiers & Lindblad (2002) Hydrogenases and Hydrogen Metabolism of Cyanobacteria. *Microbiology and Molecular Biology Reviews* 66:1-20
- Tamagnini, Leitão, Oliveira, Ferreira, Pinto, Harris, Heidorn & Lindblad (2007) Cyanobacterial hydrogenases: diversity, regulation and applications. *FEMS Microbiology Reviews* 31:692-720
- Truong & Ikura (2001) The use of FRET imaging microscopy to detect protein–protein interactions and protein conformational changes *in vivo*. *Current Opinion in Structural Biology* 11:573-578
- Wei, Ramasubramanian & Golden (1994) *Anabaena* sp. Strain PCC 7120 *ntcA* Gene Required for Growth on Nitrate and Heterocyst Development. *Journal of Bacteriology* 176:4473-4482
- Yoshikawa, Kojima, Nakajima, Furusawa, Hirasawa & Shimizu (2011) Reconstruction and verification of a genome-scale metabolic model for *Synechocystis* sp. PCC6803. *Appl Microbiol Biotechnol* 92:347-358
- Young (1998) Yeast Two-Hybrid: So Many Interactions, (in) So Little Time... *Biology of Reproduction* 58:302-311
- Zhang & Bryant (2011) The Tricarboxylic Acid Cycle in Cyanobacteria. *Science* 334:1551-1553
- Zheng, Szustakowski, Fortnow, Roberts & Kasif (2002) Computational Identification of Operons in Microbial Genomes. *Genome Research* 12:1221-1230
- URL-1 (Zugriff: 30.01.2013)  
<http://vanted.ipk-gatersleben.de/>
- URL-2 (Zugriff: 30.01.2013)  
[http://www.sbgm.org/Main\\_Page](http://www.sbgm.org/Main_Page)
- URL-3 (Zugriff: 30.01.2013)  
<http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi>
- URL-4 (Zugriff: 30.01.2013)  
<http://www.atzonline.de/Aktuell/Nachrichten/1/15528/Forschungsverbund-Solar-Biofuels-Ruhr-gegruendet.html>

URL-5 (Zugriff: 30.01.2013)

<http://www.spiegel.de/wissenschaft/technik/energietraeger-der-zukunft-bakterien-liefern-rekordmengen-an-wasserstoff-a-734657.html>

URL-6 (Zugriff: 30.01.2013)

<http://www.biotechnologie.de/BIO/Navigation/DE/Foerderung/foerderbeispiele,did=113142.html>

URL-7 (Zugriff: 30.01.2013)

<http://www.biowasserstoff.de/>

URL-8 (Zugriff: 30.01.2013)

<http://genome.kazusa.or.jp/cyanobase/Synechocystis/>

URL-9 (Zugriff: 30.01.2013)

[http://www.genome.jp/kegg-bin/show\\_organism?menu\\_type=pathway\\_maps&org=syn](http://www.genome.jp/kegg-bin/show_organism?menu_type=pathway_maps&org=syn)

URL-10 (Zugriff: 30.01.2013)

<http://wholecell.stanford.edu/>

URL-11 (Zugriff: 30.01.2013)

<http://www.genome.sk.ritsumei.ac.jp/odb2/>

URL-12 (Zugriff: 30.01.2013)

<http://genomics10.bu.edu/operons/>

URL-13 (Zugriff: 30.01.2013)

<http://www.gopubmed.org/web/gopubmed/1?WEB0ksoea2u6rcyh111100h001000j1004000202010>

URL-14 (Zugriff: 30.01.2013)

<http://www.gopubmed.org/web/gopubmed/>

URL-15 (Zugriff: 30.01.2013)

<http://string-db.org/>

URL-16 (Zugriff: 30.01.2013)

<http://www.genome.jp/kegg/>

URL-17 (Zugriff: 30.01.2013)

<http://biocyc.org/>

URL-18 (Zugriff: 30.01.2013)

<http://biocyc.org/META/organism-summary?object=META>

URL-19 (Zugriff: 30.01.2013)

<http://operondb.cbcb.umd.edu/cgi-bin/operondb/operons.cgi>

URL-20 (Zugriff: 30.01.2013)

<http://genomics10.bu.edu/operons/syn.html>

URL-21 (Zugriff: 30.01.2013)

<http://bioportal.kobic.re.kr/SynechoNET/>

URL-22 (Zugriff: 30.01.2013)

<http://nucleix.mbu.iisc.ernet.in/prombase/index.htm>

URL-23 (Zugriff: 30.01.2013)

<http://nucleix.mbu.iisc.ernet.in/prompredict/prompredict.html>

URL-24 (Zugriff: 30.01.2013)

<http://www.chem.qmul.ac.uk/iubmb/enzyme/>

URL-25 (Zugriff: 30.01.2013)

<https://www.soscisurvey.de/>

URL-26 (Zugriff: 30.01.2013)

<http://de-de.facebook.com/>

URL-27 (Zugriff: 30.01.2013)

<http://genome.kazusa.or.jp/cyanobase/Anabaena>

URL-28 (Zugriff: 30.01.2013)

<http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?name=cyano>  
bacteria

URL-29 (Zugriff: 30.01.2013)

<http://www.ipk-gatersleben.de/querschnittsbereiche/bioinformatik/>

URL-30 (Zugriff: 30.01.2013)

<https://www.virtualbox.org/>

URL-31 (Zugriff: 30.01.2013)

<http://www.bina.unipax.info/features>

URL-32 (Zugriff: 30.01.2013)

<http://www.biotapestry.org/>

URL-33 (Zugriff: 30.01.2013)

<http://www.biouml.org/>

URL-34 (Zugriff: 30.01.2013)

<http://www.celldesigner.org/>

URL-35 (Zugriff: 30.01.2013)

<http://www.cytoscape.org/>

URL-36 (Zugriff: 30.01.2013)

<http://mapman.gabipd.org/web/guest/mapman>

URL-37 (Zugriff: 30.01.2013)

<http://www.13cflux.net/omix/>

URL-38 (Zugriff: 30.01.2013)

<http://www.ondex.org/>

URL-39 (Zugriff: 30.01.2013)

<http://www.pathvisio.org/>

URL-40 (Zugriff: 30.01.2013)

<http://pl.csl.sri.com/>

URL-41 (Zugriff: 30.01.2013)

<http://innetics.com/>

URL-42 (Zugriff: 30.01.2013)

<http://www.g-language.org/PathwayProjector/>

URL-43 (Zugriff: 30.01.2013)

<http://visant.bu.edu/>

URL-44 (Zugriff: 30.01.2013)

<https://simtk.org/home/wholecell>

URL-45 (Zugriff: 30.01.2013)

<http://www.ebi.ac.uk/biomodels-main/>

## ANLAGEN

Die Anlagen umfassen tabellarisch aufgearbeitet alle in den Modellen verarbeitete Gene für den Wasserstoffmetabolismus von *Synechocystis* sp. PCC 6803 und *Nostoc* sp. PCC 7120. Außerdem sind mögliche putative Promoter-Bindestellen angeführt.

**Tabelle 13: Datensatz im metabolischen Modell involvierter Proteine und die zugehörigen ermittelten Promoter-Bindestellen und deren Reliabilität mittels PromBase von *Synechocystis* sp. PCC 6803**

ORF	Name	Promoterregion (bp)		Länge (nt)	Level	Reliabilität
		Start	Ende			
<i>sll1226</i>	<i>hoxH</i>	1673630	1673505	126	5	am höchsten
<i>sll1224</i>	<i>hoxY</i>					
<i>sll1223</i>	<i>hoxU</i>					
<i>sll1221</i>	<i>hoxF</i>	1678038	1677935	104	4	sehr hoch
<i>sll1220</i>	<i>hoxE</i>	1678898	1678786	112	3	hoch
<i>sll1626</i>	<i>lexA</i>	1319421	1319305	117	5	am höchsten
<i>slr1876</i>	<i>hoxW</i>	1227020	1227035	16	4	sehr hoch
<i>sll0359</i>	<i>sll0359</i>	2146508	2146426	83	3	hoch
<i>sll0822</i>	<i>sll0822</i>	2862005	2862394	89	3	hoch
<i>slr1498</i>	<i>hypD</i>	54457	54510	54	3	hoch
		54392	54403	12	2	durchschnittlich
		54199	54244	46	2	durchschnittlich
<i>sll1079</i>	<i>hypB2</i>					
<i>sll1078</i>	<i>hypA2</i>					
<i>sll1462</i>	<i>hypE</i>					
<i>ssl3580</i>	<i>hypC</i>					
<i>sll1432</i>	<i>hypB1</i>	1908742	1908683	60	3	hoch
<i>slr1675</i>	<i>hypA1</i>	1955107	1955207	101	3	hoch
<i>sll0322</i>	<i>hypF</i>	2434817	2434779	39	4	sehr hoch
<i>sll0593</i>	<i>glk</i>	2946268	2946177	92	2	durchschnittlich
<i>slr0329</i>	<i>xyIR</i>	2283508	2283516	9	1	niedrig
<i>slr1349</i>	<i>pgi</i>					
<i>sll0745</i>	<i>pfkA</i>	3126265	3126215	51	2	durchschnittlich
<i>sll1196</i>	<i>pfkA</i>	295068	294961	108	3	hoch
<i>slr0952</i>	<i>fbp</i>	2021976	2022035	60	3	hoch
<i>slr2094</i>	<i>glpX</i>	1560313	1560391	79	3	hoch
<i>sll0018</i>	<i>cbbA</i>	2471740	2471685	56	3	hoch
<i>slr0943</i>	<i>fda</i>	2007808	2007862	55	3	hoch

<b>slr0884</b>	<i>gap1</i>	1146747	1146761	15	2	durchschnittlich
		1146808	1146845	38	2	durchschnittlich
<b>slI1342</b>	<i>gap2</i>	3279228	3279135	94	3	hoch
<b>slr0394</b>	<i>pgk</i>	2139113	2139207	95	3	hoch
<b>slr1124</b>	<i>gpmB</i>					
<b>slr1748</b>	<i>slr1748</i>	1011845	1011940	96	2	durchschnittlich
<b>slr1945</b>	<i>yibO</i>	2248507	2248525	19	3	hoch
		2248265	2248345	81	2	durchschnittlich
<b>slr0752</b>	<i>eno</i>					
<b>slI0587</b>	<i>pykF</i>	2959733	2959723	11	2	durchschnittlich
<b>slI1275</b>	<i>pykF</i>	1104096	1104089	8	2	durchschnittlich
<b>slI0741</b>	<i>nifJ</i>	3132153	3132012	142	3	hoch
		3132372	3132353	20	2	durchschnittlich
<b>slr0721</b>	<i>me</i>					
<b>slI0920</b>	<i>ppc</i>	1997483	1997464	20	2	durchschnittlich
		1997749	1997686	64	2	durchschnittlich
<b>slI0891</b>	<i>citH</i>	2796740	2796688	53	2	durchschnittlich
<b>slr0018</b>	<i>fumC</i>	2487926	2487987	62	2	durchschnittlich
<b>slI0823</b>	<i>sdhB</i>					
<b>slI1625</b>	<i>sdhB</i>	1320846	1320842	5	1	niedrig
<b>slr1233</b>	<i>sdhA</i>	1040349	1040468	120	5	am höchsten
<b>slI1023</b>	<i>sucC</i>	393147	393057	91	5	am höchsten
<b>slI1557</b>	<i>sucD</i>					
<b>slI0401</b>	<i>gltA</i>	2561387	2561303	85	1	niedrig
<b>slr0665</b>	<i>slr0665</i>	3348397	3348533	137	2	durchschnittlich
<b>slr1289</b>	<i>icd</i>	282878	282981	104	2	durchschnittlich
<b>slI0603</b>	<i>menD</i>	2672308	2672301	8	2	durchschnittlich
<b>slr0370</b>	<i>gabD</i>					
<b>slr1329</b>	<i>atpB</i>	1666869	1666959	91	2	durchschnittlich
<b>slr1330</b>	<i>atpC</i>	1668479	1668556	78	3	hoch
<b>slI1327</b>	<i>atpC</i>	172912	172864	49	3	hoch
		172956	172935	22	3	hoch
<b>slI1326</b>	<i>atpA</i>	174577	174505	73	3	hoch
<b>slI1325</b>	<i>atpD</i>	175591	175537	55	2	durchschnittlich
<b>slI1324</b>	<i>atpF</i>	175708	175701	8	2	durchschnittlich
		176106	176101	6	2	durchschnittlich
<b>slI1323</b>	<i>atpG</i>	176586	176511	76	3	hoch
<b>ssl2615</b>	<i>atpH</i>	176586	176511	76	3	hoch
<b>slI1322</b>	<i>atpI</i>					
<b>slr1239</b>	<i>pntA</i>	1053821	1053910	90	2	durchschnittlich
<b>slr1434</b>	<i>pntB</i>	1056894	1057013	120	5	am höchsten
		1056710	1056730	21	5	am höchsten
<b>slr1843</b>	<i>zwf</i>	961054	961159	106	3	hoch

<b><i>sll1479</i></b>	<i>devB</i>	3395755	3395685	71	4	sehr hoch
<b><i>sll0329</i></b>	<i>gnd</i>	2389160	2389069	92	2	durchschnittlich
<b><i>sll0807</i></b>	<i>cfxE</i>	1713370	1713257	114	3	hoch
<b><i>sll1070</i></b>	<i>tkrA</i>					
<b><i>slr1793</i></b>	<i>talB</i>					
<b><i>slr0194</i></b>	<i>rpiA</i>	2725220	2725334	115	4	sehr hoch
<b><i>ssl2153</i></b>	<i>rpiB</i>	1923318	1923256	63	2	durchschnittlich
		1923380	1923369	12	2	durchschnittlich
<b><i>sll1525</i></b>	<i>prk</i>	2067263	2067216	48	1	niedrig
<b><i>slr0009</i></b>	<i>rbcL</i>	2478296	2478310	15	2	durchschnittlich
		2478171	2478231	61	2	durchschnittlich
<b><i>slr0012</i></b>	<i>rbcS</i>	2480186	2480207	22	3	hoch

**Tabelle 14: Datensatz im metabolischen Modell involvierter Proteine und die zugehörigen ermittelten Promoter-Bindestellen und deren Reliabilität mittels PromBase von *Nostoc* sp. PCC 7120**

ORF	Name	Promoterregion (bp)		Länge (nt)	Level	Reliabilität
		Start	Ende			
<b><i>alr0751</i></b>	<i>hoxE</i>					
<b><i>alr0752</i></b>	<i>hoxF</i>	873273	873365	93	5	am höchsten
<b><i>alr0762</i></b>	<i>hoxU</i>					
<b><i>alr0764</i></b>	<i>hoxY</i>					
<b><i>alr0766</i></b>	<i>hoxH</i>	886320	886367	48	3	hoch
<b><i>alr4908</i></b>	<i>lexA</i>	5854191	5854303	113	4	sehr hoch
<b><i>all0770</i></b>	<i>hoxW</i>					
<b><i>alr0694</i></b>	<i>hypF</i>					
<b><i>asr0695</i></b>	<i>hypC</i>					
<b><i>alr0946</i></b>	<i>calA</i>					
<b><i>alr0696</i></b>	<i>hypD</i>					
<b><i>alr0698</i></b>	<i>hupE</i>	804835	804903	69	3	hoch
<b><i>alr0699</i></b>	<i>hupA</i>					
<b><i>alr0700</i></b>	<i>hupB</i>					
<b><i>all0688</i></b>	<i>hupS</i>	796557	796536	22	2	durchschnittlich
		796841	796717	125	4	sehr hoch
<b><i>all0687</i></b>	<i>hupL</i>	795506	795486	21	2	durchschnittlich
<b><i>alr1423</i></b>	<i>hupW</i>	1686674	1686705	32	2	durchschnittlich
<b><i>alr0677</i></b>	<i>xisC</i>	785588	785608	21	4	sehr hoch
<b><i>all1440</i></b>	<i>nifK</i>	1700713	1700653	61	4	sehr hoch
<b><i>all1454</i></b>	<i>nifD</i>					
<b><i>all1455</i></b>	<i>nifH</i>	1714537	1714455	83	2	durchschnittlich



<b>alr2973</b>	<i>glk</i>					
<b>alr1982</b>	<i>alr1982</i>	2370319	2370437	119	4	sehr hoch
<b>all5002</b>	<i>all5002</i>	5975884	5975839	46	2	durchschnittlich
		5976079	5976071	9	2	durchschnittlich
<b>all1371</b>	<i>all1371</i>	1625975	1625840	136	4	sehr hoch
<b>alr1050</b>	<i>pgi</i>	1220308	1220378	71	2	durchschnittlich
		1220173	1220184	12	2	durchschnittlich
<b>alr1913</b>	<i>pfkA</i>	2292987	2293091	105	5	am höchsten
<b>all4021</b>	<i>fbp</i>	4843210	4843209	2	2	durchschnittlich
<b>alr1041</b>	<i>glpX</i>					
<b>all3735</b>	<i>all3735</i>	4511854	4511849	6	1	niedrig
<b>all4563</b>	<i>fda</i>	5458757	5458624	134	4	sehr hoch
		5459075	5459070	6	4	sehr hoch
<b>all5062</b>	<i>gap2</i>	6035671	6035606	66	3	hoch
<b>all2566</b>	<i>gap1</i>					
<b>alr1095</b>	<i>gap3</i>	1283053	1283183	131	4	sehr hoch
<b>all4131</b>	<i>pgk</i>					
<b>all4182</b>	<i>all4182</i>	5032041	5031919	123	4	sehr hoch
<b>all4237</b>	<i>all4237</i>	5074109	5074056	54	2	durchschnittlich
<b>alr1107</b>	<i>alr1107</i>	1295892	1295912	21	1	niedrig
<b>alr3338</b>	<i>alr3338</i>					
<b>all3538</b>	<i>eno</i>	4264488	4264368	121	3	hoch
<b>all2564</b>	<i>pyk1</i>	3065491	3065465	27	2	durchschnittlich
<b>all4008</b>	<i>all4008</i>	4828963	4828917	47	4	sehr hoch
		4828999	4828986	14	5	am höchsten
		4829365	4829322	44	2	durchschnittlich
<b>alr1911</b>	<i>nifJ</i>	2287954	2288086	133	2	durchschnittlich
		2287629	2287724	96	5	am höchsten
<b>alr2803</b>	<i>nifJ</i>	3409558	3409626	69	3	hoch
		3409434	3409440	7	1	niedrig
<b>alr4596</b>	<i>alr4596</i>	5493985	5494101	117	4	sehr hoch
<b>all4861</b>	<i>pepC</i>	5791468	5791335	134	4	sehr hoch
		5791842	5791796	47	2	durchschnittlich
<b>alr4322</b>	<i>alr4322</i>					
<b>alr0222</b>	<i>gltA</i>					
<b>all1267</b>	<i>all1267</i>	1506371	1506229	143	4	sehr hoch
<b>alr1827</b>	<i>icd</i>					
<b>alr0312</b>	<i>menD</i>	356377	356444	68	4	sehr hoch
<b>all3556</b>	<i>all3556</i>	4288598	4288580	19	3	hoch
<b>all3913</b>	<i>all3913</i>					
<b>all3914</b>	<i>all3914</i>	4719463	4719371	93	1	niedrig
<b>all0945</b>	<i>all0945</i>	1098685	1098627	59	2	durchschnittlich
<b>all2970</b>	<i>sdhA</i>	3614713	3614658	56	3	hoch

		3614756	3614736	21	5	am höchsten
		3615113	3615029	85	2	durchschnittlich
<b>alr3724</b>	<i>aspA</i>					
<b>all5039</b>	<i>atpB</i>					
<b>all0005</b>	<i>atpA</i>					
<b>all0004</b>	<i>atpC</i>	4475	4371	105	4	sehr hoch
<b>all0006</b>	<i>atpD</i>	7129	7112	18	2	durchschnittlich
<b>all5038</b>	<i>atpC</i>	6010052	6010024	29	2	durchschnittlich
<b>asl0009</b>	<i>atpH</i>	8298	8264	35	4	sehr hoch
<b>all0010</b>	<i>atpI</i>	9454	9408	47	3	hoch
<b>all0007</b>	<i>atpF</i>					
<b>all0008</b>	<i>atpG</i>	7913	7905	9	2	durchschnittlich
		8298	8264	35	4	sehr hoch
<b>all0011</b>	<i>atp1</i>					
<b>all3408</b>	<i>all3408</i>					
<b>all3410</b>	<i>all3410</i>	4121641	4121510	132	3	hoch
<b>asl3409</b>	<i>asl3409</i>					
<b>all4019</b>	<i>zwf</i>					
<b>alr1602</b>	<i>devB</i>	1883308	1883310	3	3	hoch
		1883218	1883285	68	3	hoch
		1883000	1883016	17	2	durchschnittlich
<b>alr5275</b>	<i>alr5275</i>	6294229	6294332	104	4	sehr hoch
<b>alr0782</b>	<i>alr0782</i>					
<b>all4052</b>	<i>all4052</i>	4879841	4879707	135	4	sehr hoch
<b>alr3344</b>	<i>alr3344</i>	4040319	4040418	100	3	hoch
		4040066	4040103	38	3	hoch
<b>all2563</b>	<i>tal</i>	3063565	3063493	73	3	hoch
<b>all4020</b>	<i>tal</i>					
<b>all0888</b>	<i>all0888</i>	1027599	1027586	14	2	durchschnittlich
<b>alr2350</b>	<i>alr2350</i>	2832812	2832933	122	3	hoch
<b>alr4123</b>	<i>prk</i>					
<b>alr1524</b>	<i>rbcL</i>	1785932	1786039	108	4	sehr hoch
<b>alr1526</b>	<i>rbcS</i>					
<b>all4312</b>	<i>nrrA</i>	5167963	5167880	84	5	am höchsten
		5167989	5167986	4	1	niedrig

**Tabelle 15: Darstellung der weiteren Antworten bezüglich der Ansprüche an einen Visualisierungsstandard**

4. Stellen Sie explizite Ansprüche an einen Visualisierungsstandard? Gern können Sie nachfolgend einige Ihrer Wünsche äußern.
Ein Standard sollte eine Sache nicht verkomplizieren, sondern sie vereinfachen.
Durchsetzung könnte problematisch sein, solange nicht eine Autorität wie z.B. IUPAC dahintersteht
keine Kenntnisse zur Software
selbsterklärend
Annotation mit EMBL und NCBI gleich
leicht zu erlernen
Ich finde die Darstellung im obigen Beispiel graphisch nicht sehr überzeugend, zu "wild". Hier würde ich mir eine übersichtlichere Form wünschen.
Yes, but no requests at the moment.
It would be helpful to connect SBGN to a mathematical representation (e.g. SBML). Perhaps implement SBGN as an extension to SBML so the graph can be specified alongside the math.
It would be helpful to include provisions for plotting experimental data.
Connections to graph layout algorithms would be helpful.
SBGN erlaubt keine kurvig- oder kreisförmig angeordneten Stoffwechselwege. Die erlaubt nicht das Einbinden von Strukturformeln.
All das ist jedoch defacto-Standard unter Biologen. Ein Blick in Fachbücher oder Google-Bildersuche "Zentralstoffwechsel" genügt um das zu erkennen.
Ich bin überzeugt, dass sich SBGN außerhalb der Bioinformatik nicht in der beabsichtigten Breite etablieren wird.
Abgesehen davon ist es einer unter vielen Versuchen, eine standardisierte Notation einzuführen.
Man kann nicht argumentieren, dass man das Problem der Interpretierbarkeit graphischer Symbole gelöst hat, indem man einen Satz neuer Symbole einführt und ihn als "Standard" postuliert.
Im Übrigen finde ich diese Umfrage sehr populistisch. Umfragen bestehen aus neutral gehaltenen Fragen und nicht aus einem Werbetext.

## EIGENSTÄNDIGKEITSERKLÄRUNG

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe. Stellen, die wörtlich oder sinngemäß aus Quellen entnommen wurden, sind als solche kenntlich gemacht. Diese Arbeit wurde in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegt.

Rochlitz, 2013



---

Eric Frenzel